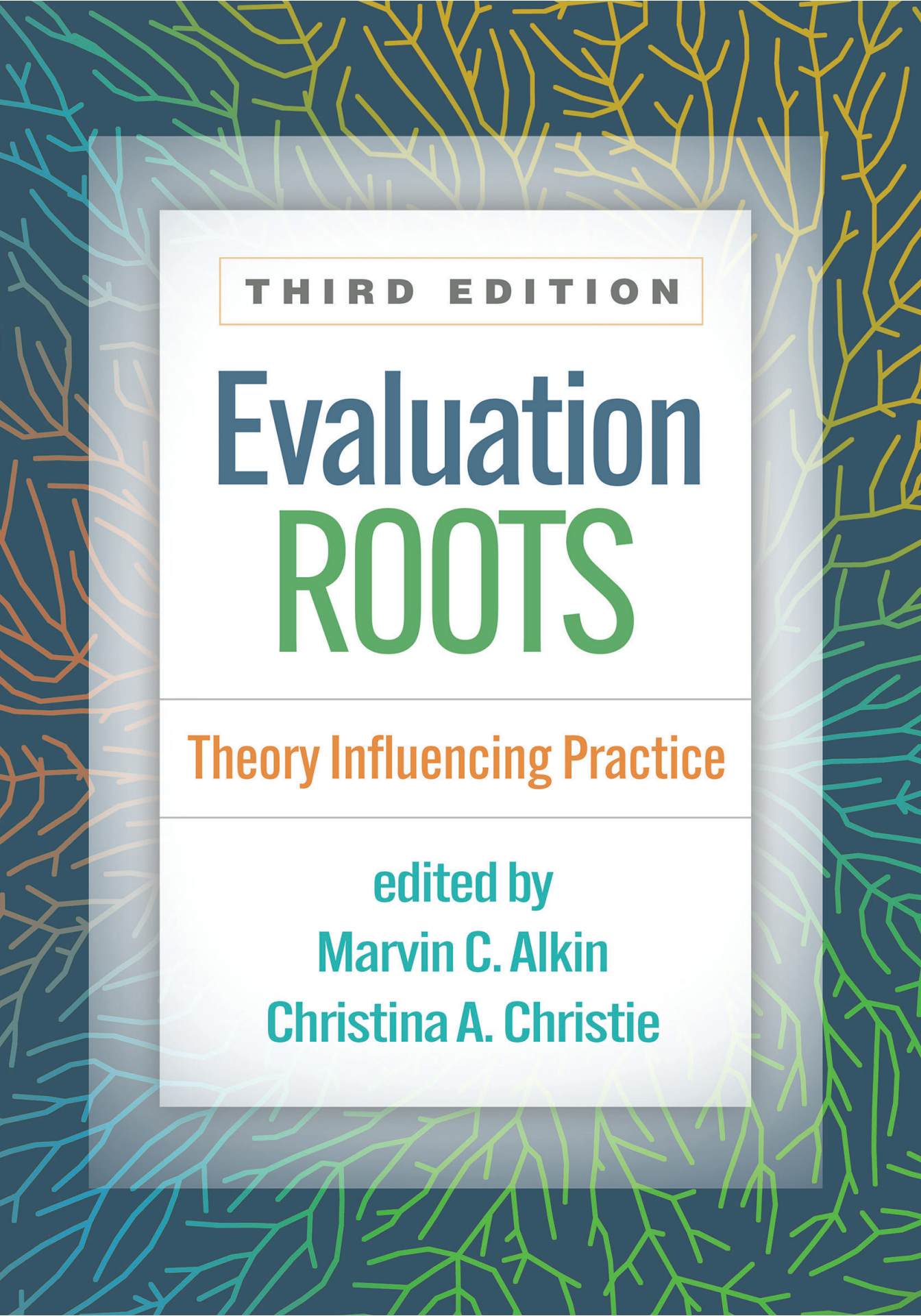




THIRD EDITION

Evaluation ROOTS

Theory Influencing Practice

edited by
Marvin C. Alkin
Christina A. Christie



GUILFORD PRESS

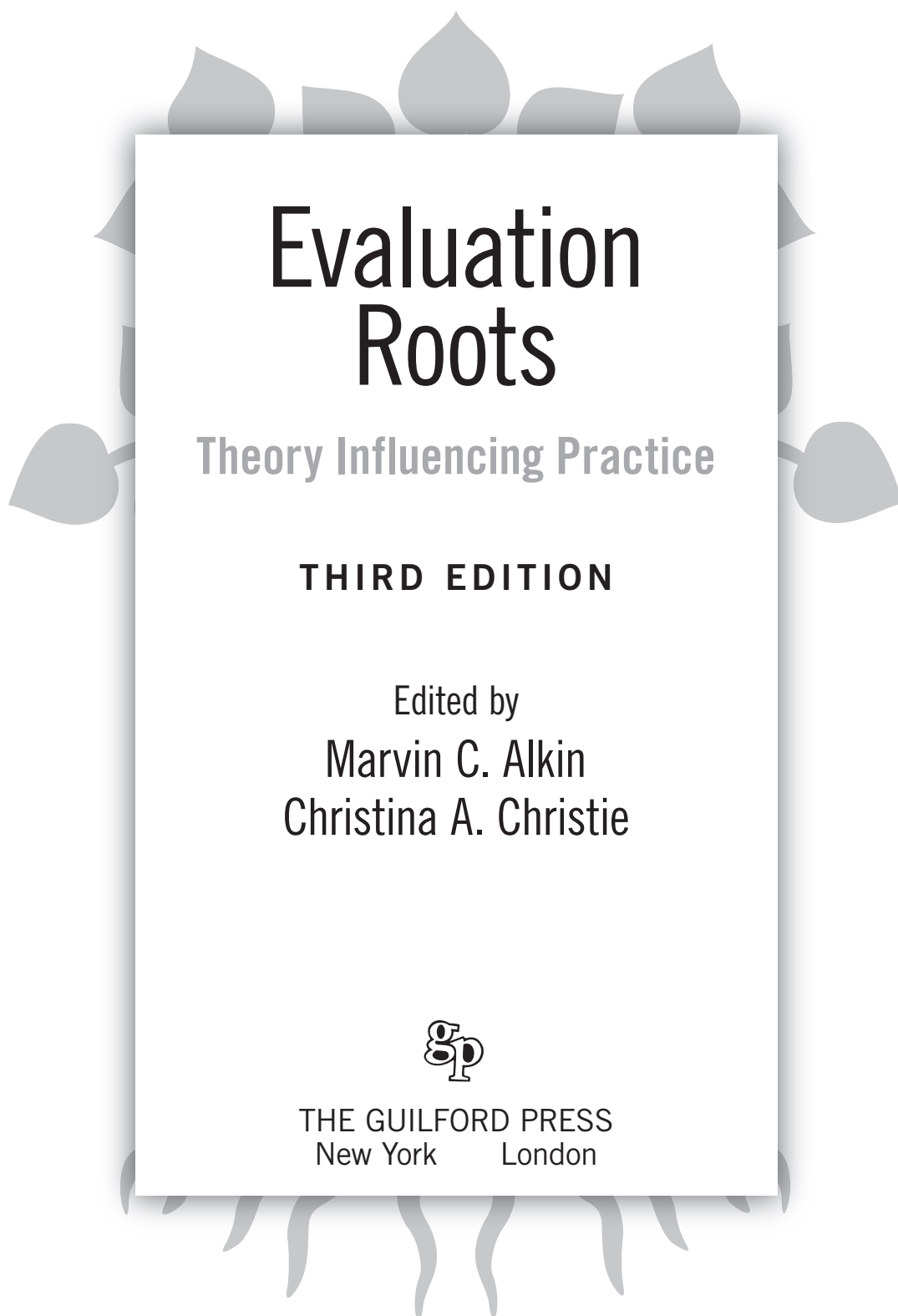
e-book

EVALUATION ROOTS

Also Available

Evaluation Essentials, Second Edition:
From A to Z

Marvin C. Alkin and Anne T. Vo

A decorative graphic of stylized tree roots and leaves in a light gray color, framing the central white text area. The roots are at the bottom, and the leaves are at the top and sides.

Evaluation Roots

Theory Influencing Practice

THIRD EDITION

Edited by
Marvin C. Alkin
Christina A. Christie



THE GUILFORD PRESS
New York London

Copyright © 2023 The Guilford Press
A Division of Guilford Publications, Inc.
370 Seventh Avenue, Suite 1200, New York, NY 10001
www.guilford.com

All rights reserved

No part of this book may be reproduced, translated, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the publisher.

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data is available from the publisher.

ISBN: 978-1-4625-5139-2 (paperback)

ISBN: 978-1-4625-5140-8 (cloth)

Editors' Introduction

In *Evaluation Roots: Theory Influencing Practice, Third Edition*, we examine and compare current evaluation theoretical stances. The point of view of the book is that there is a relationship between evaluation theories, in that theories proposed generally build on previous theoretical work. Moreover, we consider the relationship between social science research methodology, social accountability, and epistemology. Seeing the way in which these evaluation “roots” grew over time into an evaluation theory tree, with different limbs for different evaluation theory orientations, helps provide a better understanding of not only evaluation theory but also how an evaluator may go about picking the most appropriate and applicable theory for his or her evaluation.

In the three introductory chapters, we present an evaluation theory tree with three major limbs and review the way in which this edition was developed. We posit that evaluation theories are distinguished by the extent to which they place greatest emphasis on methods, valuing, or use. The extent to which a theory places highest priority on one or another of these dimensions determines its placement on a particular limb of the evaluation theory tree. Following these first three chapters are 18 chapters written by theorist authors to provide insight into the nature of their theory and how to apply it to an evaluation, and a further chapter provides a synopsis of several theories deemed important but for which theorists were unable to write chapters. Then there are two final chapters intended to support evaluators in using these theories as a basis for influencing evaluation practice and evaluation policy.

The chapter authors were requested specifically to address their respective theories and provide a holistic view of how these theories could be applied to specific evaluations. While there was some editorial advice given on occasion, chapters were ultimately accepted in the style used by the authors. As a result, writing styles differ from chapter to chapter. Some are very formal (akin to research papers), and others are more personal, capturing further insights about each author and his or her experiences. We view not only the author's description of the evaluation theory but also his or

her presentation style as a basis for providing greater perspective on the author. That being said, though, through whichever means each author saw fit, the chapters were written specifically to support evaluators in implementing effective evaluations.

The process of editing this book and reading all the chapters to provide comments to their authors has in many ways been confusing. Chapter authors present an abundance of appealing ideas, so that one constantly finds oneself saying, "You're right." Each theorist may believe that his or her approach is the true nirvana. While we have theoretical prescriptive preferences, we nonetheless believe that there is validity in the models and that each may be helpful for specific contexts. In examining their views, readers and practitioners would be advised to use not only the models that make the most intellectual sense to them but also the ones that are most fitting for their setting and goals for the evaluation. We have provided assistance for this endeavor by developing paradigms for each theory, to be found in Chapter 23. It may be that some practitioners are faced with a wide variety of different contexts in which they will perform an evaluation and select the theoretical approach that best fits each situation.

We anticipate that in reading this book, practitioners and theorists will find many appealing ideas to enhance their evaluation views and practice.

ACKNOWLEDGMENTS

Our sincere thanks to the evaluation scholars who wrote chapters for this edition of *Evaluation Roots*. We are gratified by their willingness to participate in this project and to produce engaging chapters that provide insight into the practical applications of evaluation theory.

We also thank the editors of The Guilford Press for publishing this book and engaging with us in its production. As with each of Marvin C. Alkin's other books, it has also been a joy working with C. Deborah Laughton. We are grateful for all her assistance and engagement that has moved this book from an idea to a reality. We are also so very grateful for the work, support, and contributions of Naomi Stephen. She has been a brilliant partner in the production of this book.

Contents

PART I. INTRODUCTION

1. Evaluation Roots: Theory to Practice	3
Christina A. Christie, Jennifer E. Ho, and Marvin C. Alkin	
2. Evaluation Theory	6
Marvin C. Alkin	
3. An Evaluation Theory Tree: A Framework for Guiding a Range of Evaluation Practice	12
Christina A. Christie and Marvin C. Alkin	

PART II. METHODS

4. Randomized Field Experiments: Advances in Practice	55
Robert F. Boruch and Herbert M. Turner, III	
5. Multiple Paths to Evaluation Influence and Social Betterment	67
Melvin M. Mark and Gary T. Henry	
6. Theory-Driven Evaluation Approaches: Conventional and Innovative	78
Huey T. Chen	

PART III. VALUING

7. Valuing Science: Theory as a Guide to Practice	93
Michael Scriven	
8. Responsive Evaluation	100
Robert E. Stake and Merel Visse	
9. Deliberative Democratic Evaluation: The Evaluator as Moral Fiduciary	110
Ernest R. House	
10. Values-Engaged Democratic Evaluation: Evaluation and Democracy	118
Jennifer C. Greene	
11. Culturally Responsive Evaluation: Critical and Liberatory Roots	125
Rodney Hopson and Vidhya Shanker	
12. A Transformative Lens in Evaluation: A Pathway to Transformative Change	135
Donna M. Mertens	
13. Culturally Responsive Indigenous Evaluation: An Indigenous Journey through Culturally Responsive Approaches	145
Fiona Cram and Jill Anne Chouinard	

PART IV. USE

14. Practical Policy Evaluation: Aligning Theory with Practice	161
Eleanor Chelimsky	
15. Context-Sensitive Evaluation: An Approach to Research-Based Use	173
Marvin C. Alkin	
16. The Ongoing Evolution of Utilization-Focused Evaluation: Practice Informed by Theory, Research, and Reflection	183
Michael Quinn Patton	
17. Empowerment Evaluation: It Works in Practice, but Does It Work in Theory?	193
David M. Fetterman	

- 18.** The Curious Case of Practical Participatory Evaluation:
Explorations of the Theory–Practice Relationship 211
J. Bradley Cousins and Hind Al Hudib
- 19.** Learning-Centered Evaluation: The Transformational Power
of Evaluation Redux 225
Hallie Preskill
- 20.** The “Theory” of Interactive Evaluation Practice 234
Jean A. King
- 21.** Developmental Evaluation 243
Michael Quinn Patton

PART V. OTHER THEORIES

- 22.** Evaluation Theory Synopses 255
Alana R. Kinarsky and Emi Fujita-Conrads

PART VI. CONCLUSION

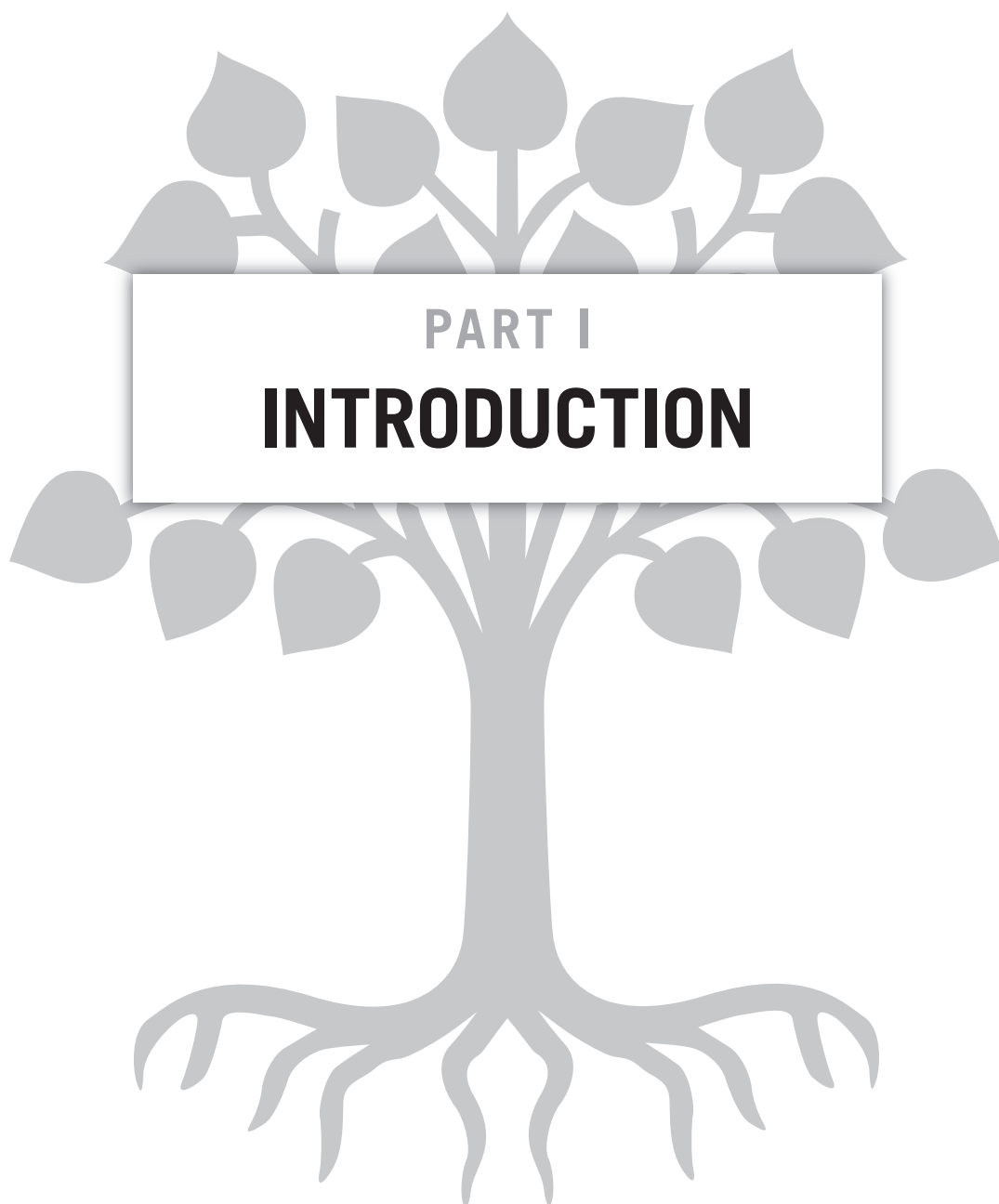
- 23.** Theory Influencing Practice 273
Marvin C. Alkin, Christina A. Christie, and Naomi A. Stephen
- 24.** Evaluation Policy as a Bridge between Evaluation Theory
and Practice: The Office of Indian Health Services, a Case Example 292
Emi Fujita-Conrads, Christina A. Christie, and Elizabeth FastHorse

Author Index 301

Subject Index 305

About the Editors 314

Contributors 315



PART I

INTRODUCTION

CHAPTER 1

Evaluation Roots Theory to Practice

Christina A. Christie
Jennifer E. Ho
Marvin C. Alkin

The first publication of *Evaluation Roots* in 2002 represented an initial effort to bring together literature from a variety of domains, sectors, and schools of thought that, in concert, represented a theoretical approach to our understanding of evaluation as a field of practice. Rather than assert any one methodology or set of principles as a primary definition of evaluation, *Evaluation Roots* reached back in history to detail the many approaches articulated by published and prolific thought leaders influencing modern evaluation practice. The compendium primarily sought to contextualize the state of the field in view of the early ideas of evaluation and “to see the way they were enhanced over time.”

In its time, *Evaluation Roots* became a scholarly centerpiece for a field that was itself experiencing a period of rapid growth. Evaluation in the 1960s and 1970s was born in the United States from The Great Society era, where federal expansion and investment in social welfare programming gave rise to

experimental examinations of how public dollars should be allocated and whether they were effective in producing social good. From the 1970s to the 1990s, while large, federally funded projects declined, evaluation was increasingly applied to a growing variety of smaller-scale programs and interventions in both the public and private sectors. Over these 30 years, evaluation became an established field through professionalization (e.g., in the identification of evaluator competencies and the growth of evaluation professional associations) and institutionalization (e.g., in embeddedness of evaluation roles within both implementing and funding organizations). The year 2000 might be characterized as the outset of an “evidence-based society,” where the value for program and policy evaluation grew increasingly dependent on the rigorous and systematic gathering, analysis, and interpretation of evidence to inform decision making. The use of evaluation in identifying “best practices”—not just effec-

tive, decent, or better practices, but “best” (Patton, 2001, p. 330)—not only demanded the systematic review of well-defined bases of evidence but also focused on increasingly widespread and diversified data sources (e.g., Big Data). It also was incumbent on the evaluation field in this period to assert the rigor of those approaches applied by evaluators in accumulating and generalizing bodies of knowledge.

Within this time line, *Evaluation Roots* was an approach to gathering together early documentation of how evaluation was or should be designed, implemented, and reviewed. The authors focused on the subject of evaluation theories, thereby introducing the novel idea that evaluation is substantiated by a theoretical base. What could perhaps be more colloquially described as evaluation approaches or models, the presented evaluation theories culminated in a distilling of individuals who were “prominent” in evaluation and “contributed substantially” to basic research methodology, evaluation issue analysis, and evaluation interpretation and teaching; and who have also advocated for a particular theoretical position on evaluation. The object of the compendium was not to map together all evaluators and approaches intersecting with these themes. Rather, selection of featured theories was based on their address of (1) issues related to the methodology being used, (2) the manner in which data are to be judged or valued, and/or (3) the user focus of the evaluation effort.

As *Evaluation Roots* itself asserts, values are an integral component of judgment. The compilation of a theoretical base explicitly relied on published work written in the English language. Individuals most eligible for consideration were those best positioned to prolifically produce peer-reviewed publications and engage in academic scholarship. As a result, featured contributors were predominantly Western, White, and male. This representation is, as a result, mirrored within the original evaluation theory tree. While the intention behind identifying indi-

vidual theorists as the “roots” of evaluation practice was to trace historical influences on current approaches, this also implicitly attributed to a select group of thinkers the genesis of the field. Black, Indigenous, and People of Color (BIPOC) scholars, as well as global and indigenous practices, were largely missing from the picture and—many have argued—from subsequent discussions around what ideas should influence the role and implementation of evaluation.

The 2020s in the United States may well be characterized as a period of public reckoning with its history of racial injustice. The murders of George Floyd, Breonna Taylor, and so many other high-profile police killings sparked a social movement that has elevated, with a sense of urgency, the need to seek truth and to redress and reform historical and deeply systemic racial injustices. This requires both acknowledgment of previously underrecognized scholars and a reimagining of multiple social structures and systems of belief that collectively enable abuse and socioeconomic harm against BIPOC. The field of evaluation is no exception, and an examination of how evaluation practices do or do not address racial equity is warranted.

It is therefore incumbent on the evaluation community to interrogate the values we bring to bear on how we define credible evidence, rigor, and the appropriateness of applying different evaluation theories and models in current contexts. This edition of *Evaluation Roots* attempts to broaden both its portfolio of contributors and the diversity of approaches discussed. Its approach is still an examination of available literature written in the English language, and the volume has known limitations in that it is not entirely comprehensive. There is still much to be done in the field of evaluation (and academia writ large) to expand the representation and accessibility of published works from a greater diversity of perspectives and experience.

In the spirit of critical reflection, this volume of *Evaluation Roots* invites readers to

consider the connections between evaluation theory and practice rather than assert which individuals constitute the “roots.” The objectives of each evaluation approach, the types of settings and scale to which they can be applied, the ways in which key stakeholders are engaged, and the limitations of each approach are presented for reader examination and assessment. The intensive review of these aspects is intended to encourage dialogue around the values, contexts, and implications of each approach, as well as to present an opportunity to consider how our own and others’ values play out in evalua-

tion implementation. The editors believe that collective deliberation and ongoing discussion are critical to our advancement of evaluation practice in a way that encourages us to be increasingly imaginative, encouraging of diversity, and ultimately, in service of equity.

REFERENCE

- Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22(3), 329–336.

CHAPTER 2

Evaluation Theory

Marvin C. Alkin

INTRODUCTION

The dictionary would have us believe that an evaluation theory is “a formulation of apparent relationships or underlying principles of certain observed phenomena which have been verified to some degree” (Theory, 2014). That is to say, a theory typically functions to provide a research-validated description of what will occur if a particular set of actions takes place. In discussing evaluation theories, Ellett and I have called such an entity a *prescriptive theory* (Alkin & Ellett, 1985), which is a set of statements and generalizations that describe, predict, or explain evaluation activities. Such a theory is designed to offer empirical understanding.

The “theories” that we find in evaluation are not theories in the way that we think of them in the physical or natural sciences. These kinds of theories do not exist in evaluation because evaluation is an applied practice field; thus, we are challenged in developing an empirically derived understanding of what results from the conduct of an evaluation; that is, for example, we are not able to definitively say that the evalua-

tion findings will be used when a particular use-oriented theory is implemented. Henry and Mark (2003) and others have advocated for conducting research on evaluation. I applaud such efforts. But we are a long way from bringing evaluation theory under the broader “theoretical umbrella.”

Without a robust research base for understanding evaluation, we must rely on the models and approaches generated by knowledgeable members of the evaluation community to guide our evaluation practice. Thus, what we do have is a variety of so-called “theories” that prescribe how each author believes an evaluation should be conducted. These are known in the literature as prescriptive theories (Alkin & Ellett, 1985) and are guides to conducting evaluation as prescribed by particular evaluation scholars. It is the convention of the evaluation literature to refer to prescriptive approaches as theories, so we, too, refer to the models described in this book as theories, and to those who have developed these models as theorists.

Almost everyone who has written about educational evaluation has in one way or another made prescriptions. Some have

spent more time systemizing their standards, criteria, and principles. A few have tried to defend or justify their prescriptions. None of the approaches is predictive.

Scholars who have developed theories have come to their views, their prescriptions, in various ways. Some were driven by their scholarly training (Christie & Alkin, 2013). Others were influenced by theories of philosophy of science. Some theory writers drew heavily from theories of justice and principles of social justice. A few drew heavily from their own research and the research of others related to the specific goals of their evaluation theories. Others were influenced by their field of academic practice. For academics in the field of education, for example, massive funding for programs designed to focus on students with special needs or instruction in science and mathematics was a major impetus in the “discovery” of evaluation. The accompanying need for evaluations to show the impact of these various funded programs was very influential in the development of prescriptive theories. Academics concerned with measurement, testing, and school improvement led to proposing prescriptions as to how these evaluations might be best accomplished.

Similarly, key events in the study of programs and policies primarily using applied research methods also influenced the development of evaluation theories. To meet the demand created by the implementation of large-scale Great Society programs, researchers advanced a means to fill the void by providing information to guide policymakers about the new host of federal programs. While these were initially discipline-based social science inquiries, the increase in federal funding and the need for knowledge to guide stakeholders (Weiss, 1972) led to a formulation of “evaluation research” as distinctive from social science research (Alkin & King, 2016).

But there are other, less formal kinds of influence on prescriptive theory development. Many theory writers drew from the experiences they had in conducting evalu-

ations. In this instance, the type of evaluations and the size of the programs they evaluated influenced how they prescribed the conduct of evaluations. Many practitioners have also been influenced by both their own positionality and observations of systemic issues facing our communities, drawing on their own life experiences and the inequities they’ve witnessed in the world. Others, who had not been active in evaluation practice, based prescriptions on their observations and perceptions of practice. Williams (2016) attempted to capture the nature of these less formal influences through the development of a *New Directions for Evaluation* volume in which seven authors discussed the influence of early personal experience on their theories. Similarly, Michael Patton and I conducted in-depth interviews to examine the various life experiences that contributed to the development of his various theories (Alkin & Patton, 2020).

In Chapter 3 of this volume, Tina Christie and I talk about some of these evaluation prescriptive theories. To identify theorists, we drew on published literature dating back to the development of the field. We reviewed peer-reviewed, published work written in the English language. As a result, the theorists featured are predominantly Western, White, and male. By choosing to focus on published academic literature mostly from North America, the theory tree does not capture the significant impact of global and indigenous scholars and practitioners. Our intention is to present some of the prescriptive theories that have created a theoretical base for evaluation practice. Yet, we know it is not comprehensive. There remains more work to be done to document the imperative contributions of Black, Indigenous, and People of Color (BIPOC) scholars, as well as global and indigenous practices.

We have chosen to categorize the primary writing on evaluation in the following ways: evaluation-oriented research methodology, evaluation issue analysis, evaluation teaching and interpretation, and evaluation theory.

First, some evaluation writings have contributed very substantially to the basic *research methodology* that forms the essential foundation for much of the work in evaluation. Authors of such works include Donald Campbell, Julian Stanley, Thomas Cook, Matthew Miles, and Michael Huberman, as well as many others. Of these writings, we have included as theories only the works of Donald Campbell and Thomas Cook because of the unique impact of their methodological contributions to the field of evaluation.

Next, we identify another category of evaluation writings as *evaluation issue analysis*. This work—these evaluation prescriptions—do not necessarily constitute evaluation theory but have very substantially contributed to our understanding of various aspects of evaluation. Individuals in this category would include, for example, Melvin Hall, Karen Kirkhart, Joan LaFrance, Michael Morris, Thomas Schwandt, William Shadish, and many others. None of these scholars is included on the evaluation theory tree. We identify yet another category as *evaluation teaching and interpretation*. Writings in this category are the various evaluation textbooks that, while not necessarily expounding new theory, provide an important resource for teaching about evaluation and help to interpret its nuances.

Finally, we have attempted to identify as *evaluation theory* the writings that advocate a particular evaluation theoretical position. These distinctions are difficult and subjective since many scholars in this category may not have presented a full theoretical exposition. Nevertheless, they appear to have proposed a distinct evaluation orientation or approach.

One further note about our selection and inclusion of evaluation theories. We have focused on writings that are related to the evaluation field and discipline rather than those that are nested within a field. For example, those who write only about evaluating health programs or education or social welfare and who exclusively assign another disciplinary designation to their

work are generally not included. And while it is perhaps obvious, we speak only about *program* evaluation theories and not personnel evaluation, or product evaluation. Finally, we recognize that a robust and important literature on evaluation theories written by authors outside of North America has been generated and grounded in evaluation work in the United Kingdom, elsewhere in Europe, Australia, New Zealand, South Africa, Africa, Asia, and elsewhere. But for the purposes of the discussion in this book, we have generally outlined our selection of evaluation theories as those written by North American authors.

CATEGORY SYSTEMS

There have been many prior attempts to look at the ways that various theoretical perspectives relate to each other. In earlier efforts, they have taken the form of category (or classification) systems. These simplified structures have provided a way to identify a limited set of characteristics for grouping theories. Entries within a category are deemed to belong together as they are similar with respect to their characteristic or configuration of characteristics that define the category. However, in making this judgment, the categorizer is selecting from the many aspects of the approach only those considered most essential. In many ways, this is similar to an artist's creation of a caricature, portraying someone or something by focusing on (even emphasizing) its most prominent features. Some of the earliest evaluation category systems were developed by Worthen and Sanders (1973) and Popham (1975). Subsequently, category systems were developed by House (1978), Glass and Ellett (1980), Alkin and Ellett (1985), Williams (1988), Shadish, Cook, and Leviton (1991), Alkin and House (1992), and others.

Category systems can be of great value. By grouping theories within a category, frequently, a new concept is introduced or an existing concept is reinforced. Thus, for example, Stufflebeam's Context, Input, Pro-

cess, Product (CIPP) model and my Center for the Study of Evaluation (CSE) model of the early 1970s were defined as *decision management* (Worthen & Sanders, 1973) and *decision facilitation* (Popham, 1975). This provided an easily understood way for practitioners to think about and define their preferred approaches to conducting evaluation. Category systems have also aided theory writers in better understanding perceived relationships between their own work and theories described by others. Placement within categories led theory writers to question whether their views were being adequately portrayed (see Alkin, 1991). Since category systems dramatize the features presumed to be most prominent, illuminating these features and not others might provide discomfort, as some theorists might not have seen certain emphases in their own work. Moreover, theory writers' views are not fixed in time, as would seem to be implied by reading their published works. They typically change over time, and published works often lag behind these changes. Nonetheless, perceptions of theory writers' views (whether or not still held) influence the work of others. Whatever the explanations for perceived portrayals of their views, the perceptions provided by category systems may force theory writers to reconsider their views and perhaps modify them to improve the accuracy of evaluation implementation of these theories.

While earlier category systems prior to the first edition of *Roots* served evaluation well, there were limitations to these taxonomies. These category systems showed neither a historical relationship between theories nor the intellectual connections among theoretical formulations and how certain ideas served to stimulate newer perspectives. The evaluation theory tree provides some indication of these relationships amongst theories.

AN EVALUATION THEORY TREE

The guiding framework for the analyses in the *Roots* evaluation theory tree is provided by the comparative theory work of Alkin

and colleagues (Alkin & Ellett, 1985; Alkin & House, 1992). As early as 1985, Ellett and Alkin maintained that all prescriptive theories must consider (1) issues related to the methodology being used, (2) the manner in which data are to be judged or valued, and (3) the user focus of the evaluation effort. With this in mind, we represent relationships between evaluation theories in the form of a tree, with each main branch representing one dimension: use, methods, or valuing. Thus, each theory is presented on the branch of the tree that we believe represents the main emphasis among these three.

Indeed, we recognize this tree to be a drastic oversimplification of very complex relationships. By depicting theories in relationship to other evaluation theories with which they share a common perspective, we capture only one dimension of the influences on their theoretical perspectives. For example, theory writers are influenced by colleagues, including by those whose theory is classified on another branch of the tree or who are scholars in other fields. To illustrate, Carol Weiss's chapter in Alkin (2004) identified my own work as well as that of Egon Guba, Ernest House, and Michael Patton—all theorists from other branches of the tree—as influences on her work. Moreover, she indicated she had been influenced by evaluators not on the evaluation tree: Howard Freeman, Gene Glass, William Shadish, and others. Robert Stake identified Lee Cronbach, Daniel Stufflebeam, and Ralph Tyler with other evaluators, including Mike Atkin, David Hamilton, and Tom Hastings. Bradley Cousins identified Carol Weiss from another branch of the tree and other evaluators, including Jennifer Greene and Michael Huberman.

THEORY AS A GUIDE TO PRACTICE

Theory is intended to serve as a guide to practice. However, that guide is helpful only if one fully understands the theory and is able to conduct it in a reproducible form consistent with the intent of the theorist.

This is not to say that practitioners should not develop theories of their own to fit their unique circumstances. But if one is using a theory—meaning that you want to avail yourself of the best thinking of a particular theory writer—it is important that you understand what that particular theory prescribes, the goals of that theory, and the conditions under which that expert believes the theory is applicable.

Research shows that evaluators have sometimes claimed to use a specific theory yet did not implement the theory as intended. Miller and Campbell (2006), for example, searched the literature for studies that stated that an empowerment evaluation approach was used, and they found few practitioners implemented empowerment evaluation as described by Fetterman. In a similar investigation of theory-driven evaluation, Coryn, Noakes, Westine, and Schröter (2011) found similar nonconformance with that theory. Miller (2010) addressed the general issue of what it means to employ a particular theoretical approach and developed standards for judging the applicability of theories. These include the extent of their specificity, range of application, feasibility in practice, and reproducibility by others different from the theorist. Furthermore, in 2017, I applied a variation of Miller's standards to a particular evaluation theory, and while appropriate for examining a theory in practice, it was limited as a framework for explicating a theory generally (Alkin, 2017).

THE GOAL OF THIS VOLUME

For a theory to be an appropriate guide for practice, it must be sufficiently explicit, so that in an actual evaluation, a user can accurately translate that for which the theory advocates. Aspirations and platitudes do not suffice. A theory needs full guidelines. This means that the theory must very precisely specify (1) the precise goal of the theory, (2) the type of evaluation for which it is appropriate (e.g., formative, summary for-

mative, summative, adaptive), (3) the size of the program to which it is most applicable, (4) some specific prescriptions, and (5) specific observable actions that help to define the theory. In other words, a full prototype should be provided so that the theory can be “translated” into a practical, working guide for implementation. In evaluating programs, we as evaluators seek sufficient information about what should be observed in the program to verify that the program has been implemented properly. Similarly, we believe that it is necessary to fully understand what evaluator actions—that is, what the evaluator says or does—must be observed to say with confidence that a particular theory has been employed.

So that the theories in this volume are more relevant to potential readers, we have asked each evaluation theorist who has prepared a chapter for inclusion to conform to these guidelines. With this added specificity, readers will be able to determine the appropriateness of each theory to their evaluation needs and to their personal values. This will allow evaluators to selectively use existing evaluation theories to guide their work.

REFERENCES

- Alkin, M. (1991). Evaluation theory development. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th Yearbook of the National Society for the Study of Education, Part II, pp. 91–112). University of Chicago Press.
- Alkin, M. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences*. SAGE.
- Alkin, M. C. (2017). When is a theory a theory?: Case example. *Evaluation and Program Planning*, 63, 141–142.
- Alkin, M. C., & Ellett, F. (1985). Evaluation models and their development. In T. Husén & T. N. Postlethwaite (Eds.), *International encyclopedia of education: Research and studies* (Vol. 3, pp. 1760–1766). Pergamon Press.
- Alkin, M. C., & House, E. (1992). Evaluation of programs. In M. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., Vol. 2, pp. 462–467). Macmillan.

- Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568–579.
- Alkin, M. C., & Patton, M. Q. (2020). The birth and adaptation of evaluation theories. *Journal of MultiDisciplinary Evaluation*, 16(35), 1–13.
- Christie, C. A., & Alkin, M. C. (2013). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed., pp. 11–57). SAGE.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226.
- Glass, G. V., & Ellett, F. S., Jr. (1980). Evaluation research. *Annual Review Psychology*, 31, 211–228.
- Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, 97, 69–80.
- House, E. R. (1978). Assumptions underlying evaluation models. *Educational Researcher*, 7(3), 4–12.
- Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation*, 31(3), 390–399.
- Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation. *American Journal of Evaluation*, 27(3), 296–319.
- Popham, W. J. (1975). *Educational evaluation*. Prentice Hall.
- Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation: Theories of practice*. SAGE.
- Theory. (2014). In M. Agnes (Ed.), *Webster's new world college dictionary* (5th ed., p. 1377). Macmillan.
- Weiss, C. H. (Ed.). (1972). *Evaluating action programs: Readings in social action and education*. Allyn & Bacon.
- Williams, D. D. (2016). Connections, themes, and implications for the future of evaluation. *New Directions for Evaluation*, 150, 85–105.
- Williams, J. E. (1988). A numerically developed taxonomy of evaluation theory and practice. *Evaluation Review*, 13(1), 18–31.
- Worthen, B. R., & Sanders, J. R. (1973). *Educational evaluation: Theory and practice*. Charles A. Jones.

CHAPTER 3

An Evaluation Theory Tree

A Framework for Guiding a Range of Evaluation Practice

Christina A. Christie
Marvin C. Alkin

INTRODUCTION

In this chapter, we present a revised “evaluation theory tree” (Alkin, 2004, 2012). The theory tree visually depicts a perspective on the philosophical foundations of evaluation and the different branches of work that are informed by these ideas (see Figure 3.1). The tree provides a way of grouping various theories, *written by a specific and limited set of scholars*, that cluster around three fundamental foci of evaluation theory: use, methods, and valuing.

Trees grow and change. The version of the theory tree presented here represents *theories* rather than previous versions that presented *theorists*. This is because our aim in this volume is to show how evaluation practice looks when guided by theory. It should be noted that the goal is not to describe the growth and development of evaluation practice; rather, the focus is on the translation of the conceptual to the practical. We use the theory tree to show a range of practices and thus provide brief descriptions of

approaches as context for this discussion of practice. Articulated are the objectives of each approach as described by a particular scholar of evaluation. This version of the tree also presents an expanded set of theories and, specifically, those theories that are focused on the “valuing” dimension of the tree (described next)—those that more directly take up issues of justice and culture in evaluation.

The three roots of the evaluation theory tree—social accountability, systematic social inquiry, and epistemology—underpin evaluation practice; each has served as an important influence on evaluation and has supported the development of the conceptual work of the field in different but related ways. They are the “why” of evaluation. Social accountability presents an important motivation for evaluation, especially for programs funded by government entities. We conceive of social accountability not as a limiting activity but rather as a means of improving programs and society. The second root—systematic social inquiry—

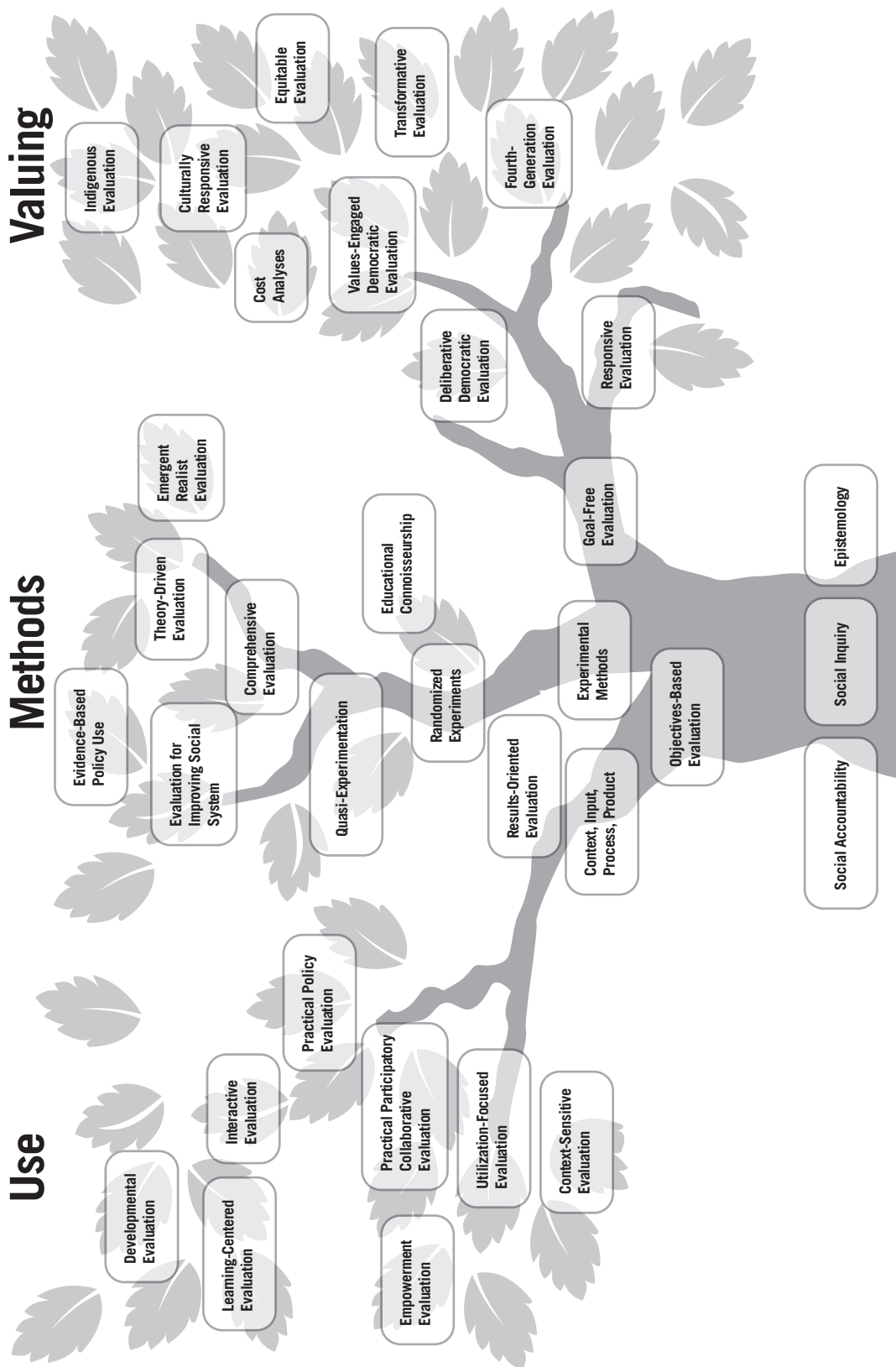


FIGURE 3.1 Evaluation theory tree.

derives from a concern for employing a methodical and justifiable set of procedures for determining accountability. Whereas accountability provides a rationale for evaluation, evaluation models are primarily derived from social inquiry. And finally, the third root, epistemology, is the area of philosophy that deals with the nature and validity (or limitations) of knowledge. Key evaluation concerns that are based in epistemological arguments include the legitimacy of value claims, the nature of universal claims, and the view that truth (or fact) is what we make it to be.

The central branch of the evaluation tree, the *methods* branch, grows from the social inquiry root. Theories on this branch are typically focused on obtaining the most rigorous knowledge possible given contextual constraints, in support of knowledge construction, as described by Shadish, Cook, and Leviton (1991). It is important to note that *rigorous knowledge* as understood in the context of this branch reflects mostly post-positivist perspectives of rigor (see Christie, Ho, & Alkin, Chapter 1, this volume, for greater detail).

To the right of the methods branch is the *valuing* branch. Valuing is an important defining activity of evaluation that distinguishes it from other applied research; valuing renders a judgment about the value and/or merit of the *evaluand*, or the subject of the evaluation. We have split this branch to reflect two fundamental perspectives—objectivist and subjectivist—that inform the valuing process. Theories grounded in *objectivist* thinking maintain that value judgments should be based on publicly observable facts. The objectivist side of the branch is closest to the methods branch because objectivist thinking is compatible with the postpositivist philosophical ideas that ground most of these theories. Alternatively, theories on the subjectivist side of the branch reflect the thought that any science that “ignores meaning and purpose is not social science” (Diesing, 1966, p. 124). Because human action is governed by sub-

jective factors, a unique characteristic of human behavior is its *subjective* meaningfulness. Theories on the right side of the valuing branch are grounded in this philosophy.

The third branch is *use*. Evaluation studies are conducted with the intent to be used. Theories on this branch have an explicit concern for how evaluation information will be used. Use is understood to be multidimensional, ranging from very specific decisions to organizational learning and evaluation capacity building of program staff and organizations. Notions about whose information and decision-making needs are prioritized in efforts to promote use are important here, as is stakeholder involvement in the evaluation process.

These three tree branches are not meant to be viewed as independent from one another; rather, they have been drawn to reflect the relational qualities among them. Thus, if the tree were in a three-dimensional space, the far right of the valuing branch would relate to (or touch) the far-left side of the use branch. Thus, the relational nature of the three branches helps to further depict approaches.

The branches of the evaluation tree relate more specifically in the following ways. Theories are positioned on each branch in a way that reflects not only their primary emphasis but also their secondary emphasis. To the immediate right of the methods branch is the objectivist arm of the valuing branch. Thus, theories placed on the right side of the methods branch, leaning toward the valuing branch, reflect a secondary importance placed on valuing. Theories on the left side of the valuing branch, which is closest to the methods branch, argue primarily for objectivist procedures for valuing, with a secondary concern for methods. For example, *cost analyses* is placed on the left side of the valuing branch to reflect the importance of methodology to this approach.

We place theories with a secondary interest in use at the far-right side of the tree—the subjectivist arm of the valuing branch. In doing so, we show the relationship between

valuing and use. These theories are more concerned with individual stakeholders' actionable use of findings than those placed on the left arm of the valuing branch. Again, imagining the tree in a three-dimensional space, the far left of the use branch contains theories with a primary emphasis on use and a secondary concern for valuing—in other words, a secondary concern for social justice and empowerment. The right side of the use branch reflects the relationship between use and methods. Specifically, theories placed here have a secondary emphasis on “rigorous” methods; those on the left side of the methods branch place a greater emphasis on use than others that fall elsewhere on the methods branch.

We want to underscore that this evaluation theory tree is a visual depiction of a range of theories within the larger evaluation theory literature. It describes how only a select set of theoretical ideas have helped to develop and influence the more expansive body of evaluation theory—done in service of describing a range of practice. We do still identify theory writers of each approach and describe how their ideas have influenced the approach in essential ways. As in prior versions, we include the theorists who are most commonly associated with or who have made substantial contributions to the literature on each of the approaches represented. For example, Rossi and Weiss have made fundamental contributions to the theory-driven evaluation approach; however, Chen (1990) synthesized and advanced these ideas in substantial ways in the first text focused exclusively on the approach, *Theory-Driven Evaluations*. Thus, Chen is the theory writer associated with the theory-driven evaluation approach.

TREE ROOTS

Social Accountability

Many scholars acknowledge the important role of evaluation in providing systematic information for the purposes of account-

ing for what has been done or accomplished as a result of the implementation of a particular program or policy. Indeed, Mark, Henry, and Julnes (2000) cite accountability as one of the four purposes of evaluation. Several decades ago, Alkin (1972) defined three types of accountability: goal accountability, process accountability, and outcome accountability. *Goal accountability* examines whether reasonable and appropriate goals have been established; *process accountability* reflects whether reasonable and appropriate procedures for accomplishing those goals have been established and implemented; and *outcome accountability* refers to the extent to which established goals have been achieved. Typically, management and program operators bear responsibility for both process and outcome accountability. All three of these accountability types are reflected in the evaluation literature. Concern for the evaluator's role in valuing goals is central to goal-free evaluation; process accountability is prominent in the “process” section of the Context, Input, Process, and Product (CIPP) model; and outcome accountability is a major focus of many evaluation efforts.

Today, evaluations with a strong focus on goal accountability are typically focused on improving institutional performance. The results of these evaluations are often used in policymaking or other governmental decision making. While the tradition of process accountability is discussed and seen less now, it had an important influence on early evaluation activities. For example, the European tradition of school inspectors, now outdated, has been included in many early depictions of school evaluation. Dreaded visits from the school inspector were designed to ensure that teachers were engaged in prescribed lessons at prescribed times. A more contemporary example of process accountability is the system of accreditation employed in secondary schools, private schools, and universities. In these instances, an evaluation team with presumed expertise, frequently guided by established

process standards, visits a site to observe, account, and make a report. The result is a judgment about whether the institution is accountable and should be accredited.

The evaluation unit of the U.S. Government Accountability Office (GAO) typifies how accountability is often thought of in contemporary North American evaluation practice. The GAO is independent, nonpartisan, and often described as the congressional watchdog because it investigates how the federal government spends tax dollars (GAO, n.d.). Its mission is to “examine the use of public funds, evaluate federal programs and activities, and provide analyses, options, recommendations, and other assistance to help the Congress make effective oversight, policy, and funding decisions through financial audits and program reviews” (see *www.gao.gov*). It is important to note that this notion of accountability implies the provision of an account to an individual or group in a position of formal authority. Thus, the GAO’s activities are designed to ensure the Executive branch’s accountability—to the Congress and all Americans. Social accountability, then, situates and legitimizes evaluation as a fundamental process for generating systematic information for decision making. This can occur within the configuration of government-sponsored programs and policies or can be more localized and occur at the intervention or program level.

Social Inquiry

Social inquiry, in its broadest sense, can be characterized as the systematic study of the behavior of individuals in groups in various social settings by a variety of methods. It begins with the recognition that human actions have a unique social dimension rather than simply a natural or psychological dimension. The central overriding question is: Why do people in social groups act as they do?

In the Western World, inquiry along these lines has its origins in 17th- and 18th-century figures such as Thomas Hobbes, Montesquieu, and Jean-Jacques Rousseau. While

these theorists systematically studied and commented on social groups, their descriptions and theories were a product more of contemplation than of rigorous empirical investigation. It was not until the mid- and late 19th century, as demonstrated in the works of Karl Marx, Émile Durkheim, and Max Weber, for instance, that society and social groups began to be studied through the collection and analysis of empirical data on social groups.

A perennial question in social inquiry is which methods are appropriate for the study of society, social groups, and social life, and, more specifically, whether the methodologies of the physical sciences, broadly defined, apply to social phenomena. Classical social theorists Marx (1867/1932) and Durkheim (1895/1966), for example, made extensive use of statistics, among other data, in forming judgments regarding social life. The discipline of psychology introduced the experimental method, where the central overriding question is whether a treatment is effective in bringing about desired effects. Discussions regarding the feasibility and desirability of this methodology for the study of the social world continue to this day, giving rise to heated debates and never-ending accusations of positivism. These debates are related to and reflective of the more general question of the applicability of the methods of the physical sciences to the social sciences.

Alternatively, the discipline of anthropology has given rise to ethnographies and, more broadly, qualitative studies of the social world. The distinction between these methods and those mentioned earlier is sometimes couched in terms of the distinction between explanation and prediction on the one hand and interpretation and understanding on the other. Clifford Geertz’s (1973) classical essay “Thick Description: Toward an Interpretive Theory of Culture,” in *The Interpretation of Cultures*, epitomizes the latter approach and in part has come to define interpretive social science, where the emphasis is placed not on prediction but on meaning.

Philosophers continue to disagree about what constitutes the methods of the sciences and their potential applicability to the study of social life—for example, as described in Tashakkori and Teddlie (1998). Cutting across social science disciplines are broad philosophical and methodological questions that continue to be debated in contemporary social inquiry. Important questions include the following: What is the relationship between theory and observation? Should social scientists have a moral stance toward the individuals and groups that they study? Is such a stance appropriate? Would it compromise the researchers' objectivity? These and other questions form part of the theory and practice of social inquiry.

Each of these social inquiry influences is readily seen in the evaluation theories discussed in this chapter. The influence of experimental psychology and quantitative sociological disciplines is readily noted in the methods branch. For example, Robert Stake's (1974) responsive evaluation involves constructs commonly associated with anthropology, such as "thick description," and concerns raised within the philosophy of science have clearly influenced Michael Scriven's (e.g., 1986) thinking about evaluation as a practice and as a discipline.

Epistemology

The vast body of philosophical literature includes discussions of paradigms. A *paradigm* is "a worldview or perspective that, in the case of research and evaluation, [that] includes conceptions of methodology, purposes, assumptions, and values . . . that typically consists of an ontology (the nature of reality), an epistemology (what is knowable and who can know it), and a methodology (how one can obtain knowledge)" (Mathison, 2005, p. 289). With these philosophical underpinnings in mind, we include in the theory tree this third root, *epistemology*, to address arguments on the nature of knowledge.

Those in evaluation who have historically and are currently engaged in epistemo-

logical discussions typically draw on one of three broad areas of thinking: postpositivism, constructivism (and related thinking), or pragmatism. Thus, the basic axioms of these paradigms offer a broader frame for understanding the theoretical influences on evaluation theory. Importantly, some evaluation theories are shaped more exactly by a philosophical theory, while in others only an undercurrent of a theory can be detected. Epistemology serves as an important influence on evaluation theory more generally, albeit to varying degrees.

Views of science shifted during the 20th century, away from positivism toward postpositivism. Where positivists believe that the goal of science is to uncover the truth, postpositivists believe its goal is to measure truth, even though that goal *cannot* be obtained because all observation is fallible and has errors. Thus, a full understanding of truth can be approached but never reached. This type of realism is referred to as *critical realism*. The ideal of "approaching" truth extends to views on causality as well: It is believed that causation is observable and that over time predictors can be established, but some degree of doubt will always remain associated with the conclusion. Values and biases are noted and accounted for, yet the belief is that they can be controlled within the context of scientific inquiry. This thinking is clearly evident in Donald Campbell's experimentation in evaluation as well in other theories on the methods branch (Christie & Fleischer, 2009).

Constructivism is one element of interpretivism. It takes a relativist stance that there is no single, tangible reality that can be approximated but rather only multiple, constructed realities. Epistemologically, constructivism views subjectivity as the only reality—that is, the only way the unknown can become known is through our own individual belief systems. As such, there are multiple realities that are subjective and that change according to the "knower," who constructs their own reality using past experiences and individual contexts to perceive what is "known." In the constructivist

paradigm, the “knower” and the “known” are interrelated, unlike in positivist epistemology, where they are considered independent. This type of realism is referred to as *ontological relativism*.

Of notable distinction, inquiry is considered value-bound rather than value-free. Constructivists believe that it is better to acknowledge and consider bias than to attempt to ignore or control it. Inductive logic is the rule, which means particular instances are used to infer broader, more general laws, although generalizability is not nearly as important as local relevance. Cause and effect are thought to be impossible to distinguish from one another because relationships are bidirectional; thus, everything impacts everything else at one time.

We describe relativism here, under constructivism, though it is a distinct school of thought. Relativism consists of various theories, each of which claims that some element or aspect of experience or culture is relative to and thus dependent on another. Some relativists claim that people can understand and evaluate beliefs and behaviors only in terms of their own historical or cultural context. This is referred to as *truth relativism*, which posits that there are no absolute truths—truth is always relative to some particular frame of reference. This is, of course, similar to notions of truth put forth by constructivists.

Pragmatists embrace objectivity and subjectivity as two positions on a continuum and argue that deductive and inductive logic should be used in concert. Pragmatists do, however, move away from embracing equally the axioms of the postpositivist and constructivist paradigms. For example, pragmatists are more similar to postpositivists regarding notions of external reality, with the understanding that there is no absolute “truth” concerning reality. However, pragmatists are more in line with constructivist thought regarding their argument of multiple explanations of reality and the belief that at any given time there is one explanation that makes the most sense. In

other words, at one point in time, a single explanation of reality may be considered “truer” than another.

TREE BRANCHES

Methods

In the beginning, there was research. And positivist and postpositivist research methodologies dominated the conduct of studies. While most evaluation theorists have methodological concerns and view applied research as the genesis of program evaluation, some have been steadfast in emphasizing that orientation. In the social sciences, an emphasis on applied research depends on well-designed experimental studies and other controls. Fundamental to these theories is the early work of Donald Campbell (1957), and specifically the better-known Campbell and Stanley volume (1966), which detailed experimental and quasi-experimental designs and the conditions of these approaches. With the exception of objectives-based evaluation (Ralph Tyler, 1942), the theories on this branch are either grounded in or have developed from this work.

Objectives-Based Evaluation: Ralph Tyler

We view Ralph Tyler’s Eight-Year Study as one of the major starting points for modern program evaluation (if one can accept the 1940s as modern). Tyler’s work was far-reaching, affecting the work of many future evaluation theorists. Indeed, Madaus and Stufflebeam (1989) credited Tyler for “ideas such as the taxonomic classification of learning outcomes, the need to validate indirect measures against direct indicators of the trait of interest . . . the concept of formative evaluation, content mastery, decision-oriented evaluation, [and] criterion-referenced and objectives-referenced tests” (p. xiii). Furthermore, Tyler avowed that the curricula to be evaluated are based on the hypotheses of program staff, their best judgments regard-

ing the most effective set of procedures for attaining program outcomes. The purpose of evaluation is to validate these hypotheses.

Madaus and Stufflebeam (1989) claimed that Tyler coined the term *educational evaluation* in the 1930s to describe his procedures—the comparison of (well-stated) intended outcomes (called *objectives*) with (well-measured) actual outcomes—and this came to be called *objectives-based (or objectives-oriented) evaluation*. Although he provided a foundation for many theoretical ideas, Tyler's main focus was on the specification of objectives and measurement of outcomes. Tyler's (1942) objectives-based evaluation focuses on (1) formulating a statement of educational objectives; (2) classifying these objectives into major types; (3) defining and refining each of these types of objectives in terms of behavior; (4) identifying situations in which students can be expected to display these types of behavior; (5) selecting and trying promising methods for obtaining evidence regarding each type of objective; (6) selecting on the basis of preliminary trials the more promising appraisal methods for further development and improvement; and (7) devising means for interpreting and using the results.

Objectives-based evaluation has had a strong continuing influence on education for many decades. As a part of the expansion of Tyler's work, Bloom, Englehart, Furst, Hill, and Krathwohl (1956) developed a taxonomy of educational objectives for the cognitive domain, and Krathwohl, Bloom, and Masia (1964) developed one for the affective domain. These efforts encouraged users to invest a great deal of time and effort into objective specification.

Changed from the version of the tree published in the first edition of *Evaluation Roots*, objectives-based evaluation has been repositioned to a subbranch. This reflects our revised view that while objectives-based evaluation is heavily methods related, this approach is not a predecessor to the approaches further up on the branch. The original positioning was intended to

reflect the influence of this approach on the field of educational evaluation, which was very significant. Upon further reflection, however, we concluded that the overall influence of the approach on the methods branch specifically was less than its original positioning suggested. Thus, we have placed objectives-based evaluation as a subbranch near the base of the methods branch.

Experimental Methods: Donald Campbell

Shadish and colleagues (1991) referred to Donald Campbell as the “methodologist of the experimenting society.” And, indeed, he was. Campbell is best known for his path-breaking work on the elimination of bias in the conduct of research in field settings. In evaluation specifically, he is best known for his papers on experimental and quasi-experimental designs for social research (Campbell, 1957; Campbell & Stanley, 1966). The focus on experimental designs (and more practical quasi-experimental designs) is an attempt to “rule out many threats precluding causal inference” (Shadish et al., 1991, p. 122).

Few scholarly works published during the 20th century had as great an impact on social science research as *Experimental and Quasi-Experimental Designs for Research* (Campbell & Stanley, 1966). This publication advanced three major areas of social science research design. First, the authors explained the conditions necessary to conduct a true experimental study, highlighting randomization. Second, they articulated and explained two key notions relevant to this: They called the degree to which an experiment is properly controlled *internal validity* and referred to the degree of applicability of the results of an experiment as *external validity*. Third, they recognized that experiments are not perfect and that they should not (and cannot) be widely applied. Thus, as an alternative to the true experiment, they described, in great detail, quasi-experimental designs.

Until that manuscript was published, the social sciences had used the scientific meth-

ods common to the physical sciences to study social problems in the name of “good science.” Many still view experimental designs as the most rigorous of all research designs, the “gold standard” against which all other designs are judged. Recognized as a striking shift in thinking, Campbell and Stanley (1966) acknowledged the shortcomings of this thinking—specifically that, in many contexts, experiments are not achievable or even desirable.

As an alternative to the experiment, Campbell and Stanley (1966) described a lengthy set of quasi-experimental designs. These are adaptations of the randomized experiment that are more easily implemented in the messy world of applied social research, where it is not always practical, ethical, or possible to randomly assign persons to experimental and control groups. Generally speaking, quasi-experimental designs include some type of intervention or treatment and provide a comparison, but they lack the degree of control found in randomized experiments. Just as randomization is the essential attribute of an experiment, the absence of randomization is the defining characteristic of a quasi-experiment. The ideas put forth in Campbell and Stanley’s 1966 manuscript remarkably impacted the promotion of new ideas related to research design and are now the foundation of almost all social science research methods courses. It could be argued that the introduction of and favorable response to quasi-experimental designs helped give rise to a climate that was apparently more accepting of alternative thinking about social science research.

While it should be noted that Campbell (1975a, 1975b) also wrote papers indicating the potential appropriateness of qualitative methods as a complement to the more quantitative experimental methods, he was truly an experimentalist. Experimental methods are the foundation of the methods branch; the principles of experimental design, quasi-experimental design, and validity serve as the basis for other approaches on

this branch, as well as a counterpoint for the development of other approaches, in particular evaluation for improving social systems (Lee Cronbach), goal-free evaluation (Michael Scriven), and responsive evaluation (Robert Stake).

*Randomized Field Experiments:
Robert Boruch*

The randomized field experiment continues to be held by many as a preferred approach to addressing evaluation questions, particularly in policy circles. In the decades that followed Campbell and Stanley’s 1966 manuscript, Campbell trained and collaborated with many influential evaluation scholars. In particular, Robert Boruch was a strong advocate for randomized field experiments in evaluation, which are distinct from the quasi-experiments put forth by Campbell and Stanley. Boruch (1997) elucidated the distinction between the study designs:

Randomized field tests are also different from *quasi-experiments*. The latter research designs have the object of estimating the relative effectiveness of different treatments that have a common aim, just as randomized experiments do, but they depend on methods other than randomization to rule out the competing explanations for the treatment differences that may be uncovered. Quasi-experiments and related observational studies then attempt to approximate the results of a randomized field test. (p. 4)

With this, evaluation is likened to conventional scientific research. It involves the systematic application of rigorous randomized research designs for measuring the extent of a social problem and assessing the implementation, relative efficacy, and cost-effectiveness ratio of social intervention programs (Boruch, Snyder, & DeMoya, 2000). Boruch has been steadfast in viewing the randomized field experiment as the most effective way of obtaining the least equivocal estimate of a social program’s effects (Boruch et al., 2000).

The randomized experiment is an approach that randomly assigns individuals (or clusters of individuals) to one of two treatment groups (Boruch, McSweeney, & Soderstrom, 1978). This procedure makes certain that groups will be statistically equivalent in the absence of program-induced differences, thereby maximizing the internal validity of the evaluation. This approach also avoids the interpretive problems that affect alternative evaluation designs by ensuring a fair comparison between groups, so that a legitimate statement can be made about the role of chance in the results (Boruch et al., 2000).

Boruch (1997) claimed that any program, be it social, educational, or welfare, should be studied systematically, employing randomized controlled experimental methods to gather valid and reliable evidence. The experimental design approach can “introduce social change in a way that allows one to effectively discern the net effect of the change on important social outcomes” (Berk, Boruch, Chambers, Rossi, & Witte, 1985, p. 390). Berk and colleagues (1985) asserted that randomized field experimentation is particularly illuminating when the goal is investigating causal inference and generalizing to other subjects and settings. Given that social policy experimentation relies on drawing samples from a known population and then randomly assigning subjects to treatments, sound generalizations will follow (Berk et al., 1985).

Evidence Matters: Randomized Trials in Education Research (Mosteller & Boruch, 2002) is an edited volume that includes arguments for the increased use of randomized trials in educational research to identify effective interventions. This volume was published at the same time that the National Research Council (2002) released *Scientific Research in Education*, which also explicitly called for an increase in randomized controlled trials (RCTs) in educational research. These two publications have been cited as offering key arguments for RCTs in educational research and helped shape the way funding for edu-

cational research was awarded under the No Child Left Behind (2001) legislation.

Randomized field trials draw directly from the theoretical principles articulated in Campbell's work on experimental designs. This is unlike other colleagues and mentees of Campbell, such as Thomas Cook, who have advocated more strongly for the use of other approaches of extermination, like quasi-experimentation, or who sought to define evaluation more broadly.

Quasi-Experimentation: Thomas Cook

The randomized field experiment is seen by some social scientists and policy analysts as the most rigorous design available for answering evaluation questions. However, as Thomas Cook pointed out, this design can sometimes be impossible to implement in social and educational program evaluation studies because of contextual factors that are beyond the control of the evaluator. As mentioned earlier, Campbell and Stanley's classic *Experimental and Quasi-Experimental Designs for Research* (1966) delineated an alternative. Quasi-experimental designs are applicable to evaluation because they address the difficulties of experimental control inherent in field research—specifically, random assignment. Generally, quasi-experiments involve the administration of pre- and posttests and they compare groups, but they do not involve randomization.

Cook and Campbell (1976, 1979) extended Campbell and Stanley's initial work, offering new quasi-experimental designs and delineating some less recognized threats to internal validity (e.g., resentful demoralization). Though Campbell struggled with his position on advocating for the use of quasi-experiments—later denouncing the use of quasi-experimental designs and saying he had committed an injustice by suggesting that quasi-experimental designs were a viable alternative to classic randomized design—Cook remained a strong proponent. He continued to focus on their use and was highly influential in developing a

field of study related to quasi-experimental designs. Cook expanded on the issues related to quasi-experimentation in several key areas. With his interests in methodology, evaluation research, and social reform, his work has been specifically concerned with the contextual factors of the evaluation and the ways these factors shape the implementation of quasi-experimental studies.

Different from other methods-driven theory writers of the time, Cook also recognized the importance of involving stakeholders when determining evaluation questions to collect information useful to decision makers. He explained that a primary problem with evaluations, as he experienced them, is that evaluators do not consult decision makers or any other interested parties when forming the research questions to be addressed. This results in general program goals that are either “operationally vague or of limited importance to many potential users” (Cook & Gruder, 1978, p. 15). Consequently, he suggested that evaluators work together with stakeholders when implementing quasi-experiments to decide what the evaluation should examine. Thus, we position quasi-experimental methods, as advanced by Cook, on the methods branch of the tree, closely following Campbell’s experimental methods, and near the use branch.

Emergent Realist Evaluation: Gary Henry and Melvin Mark (with George Julnes)

Emergent realist evaluation (ERE) is described as a comprehensive evaluation model that offers reconceptualized notions of use, methods, and valuing. Gary Henry and Melvin Mark’s collaboration with George Julnes developed into “a new theory that captures the sense-making contributions from post-positivism and the sensitivity to values from constructivist traditions” (Henry, Julnes, & Mark, 1998, p. 1). Recognizing the power of evaluation within democracies, the authors argued that “social betterment, rather than the more

popular and pervasive goal of utilization, should motivate evaluation” (Mark, Henry, & Julnes, 1998, p. 19).

ERE is an evaluation methodology that (1) gives priority to the study of generative mechanisms, (2) is attentive to multiple levels of analysis, and (3) is mixed-methods appropriate. Because ERE focuses on understanding the underlying mechanisms of programs, Mark and Henry have been concerned primarily with addressing the research questions that identify which mechanisms are operating and which are not (Mark et al., 1998). This is done in an effort to identify causal linkages and enhance the generalizable knowledge base of a particular set of programs or program theories.

Mark and Henry argued that an evaluator must determine stakeholders’ values when investigating possible mechanisms to ensure the evaluation examines the program effects that are of most interest to the public and other relevant stakeholders. Henry (1996) argued that understanding the values of the various stakeholder groups, including areas of consensus and conflict, helps to promote democratic policy. ERE offers three methods for investigating stakeholder values: (1) surveying and sampling possible stakeholders; (2) utilizing qualitative methods, such as interviews and/or focus groups to determine their needs and concerns; and (3) analyzing the context of the evaluation from a broad philosophical perspective, focusing on issues such as equity, equality, and freedom. These value investigations should then be communicated to the multiple audiences of the evaluation (Mark et al., 1998).

The ERE evaluator also engages in a process of either competitive elaboration or principled discovery. *Competitive elaboration* involves ruling out alternative explanations for study findings, which includes exploring alternative program theories and threats to validity (Mark et al., 1998). This requires a preexisting body of knowledge of possible program mechanisms and, ideally, a study design that experimentally tests each rel-

evant mechanism to identify those with the greatest or least impact. This approach lends itself to quantitative methods of inquiry, such as fully randomized experiments and quasi-experiments. *Principled discovery*, on the other hand, is used when programs are evaluated before practitioners are able to develop experientially tested theories (Mark et al., 1998). Approaches to discovering program mechanisms include exploratory data analysis, graphical methods (Henry, 1995), and regression analysis.

More recently, Mark and Henry extended their theoretical work into the area of *evaluation influence* (Henry & Mark, 2003; Mark & Henry, 2004), which is “the capacity or power of persons or things to produce effects on others by intangible or direct means” (Kirkhart, 2000, p. 7). Henry and Mark (2003) described three levels at which evaluation influence can occur—individual, interpersonal, and collective. Each level is further explained by identifying specific mechanisms, measurable outcomes, and forms of influence. To our minds, this work expands their ERE work by offering further description of the use dimension of evaluation practice within the ERE approach.

Although it leans toward the valuing branch of the tree because of its emphasis on promoting social betterment, we place ERE on the methods branch following Cook’s quasi-experimentation views because of its emphasis on the use of social science methods.

Theory-Driven Evaluation: Huey-Tsyh Chen

Huey Chen (1990, 2005; Chen & Rossi, 1992) has been most influential in developing the concept and practice of theory-driven evaluation. He has coauthored several articles with Peter Rossi on the approach and has also authored a book and articles advancing the theory. Most importantly, he wrote the first comprehensive book on theory-driven evaluation. Thus, we focus on his work here.

Chen (2005) pointed out that while he acknowledges the attractiveness of con-

trolled experiments in estimating net effects through randomization, when no effect is shown in a controlled experiment, there is no indication as to whether failure is due to poorly constructed causal linkages, insufficient levels of treatment, poor implementation, and so on. Theory-driven evaluation addresses this dilemma:

We have argued for a paradigm that accepts experiments and quasi-experiments as dominant research designs, but that emphasizes that these devices should be used in conjunction with a priori knowledge and theory to build models of the treatment process and implementation system to produce evaluations that are more efficient and that yield more information about how to achieve desired effects. (Chen & Rossi, 1983, p. 300)

As an approach, theory-driven evaluation embraces the experimental paradigm; however, this must be supplemented by the development of theoretical models of social interventions: “An unfortunate consequence of this lack of attention to theory is that the outcomes of evaluation research often provide narrow and sometimes distorted understandings of programs” (Chen & Rossi, 1983, p. 284).

In theory-driven evaluation, social science theory guides the identification of potential areas for investigation. Thus, a program theory identifies the potential effects of a program related to its explicit goals. Theory-driven evaluation is also designed to identify secondary effects and unintended consequences that go beyond a program’s stated goals by using social science knowledge and theory related to the subject matter under investigation. The theories constructed in a theory-driven evaluation are not global or grand, but “plausible and defensible models of how programs can be expected to work” (Chen & Rossi, 1983, p. 285).

Theory-driven evaluation also attempts to resolve the tensions between internal and external validity raised in Campbell’s and Cronbach’s approaches. Chen argued, “We are not convinced that the trade-off prob-

lem between internal and external validity, which is so sharply portrayed by Campbell, Cronbach, or others, that dealing with one type of validity must seriously sacrifice the other types of validity” (Chen & Rossi, 1987, p. 97). Theory-driven evaluation, Chen reasoned, enhances the possibility of achieving each type of validity, and this can be accomplished by simultaneously conducting randomized experiments and studying causation within the context of constructed theories (usually through structural equation modeling).

Because of its grounding in the methodological traditions of social science research and emphasis on using social science knowledge and theory to develop program theories, theory-driven evaluation is placed on the methods branch.

Evaluation for Improving Social Systems:
Lee J. Cronbach

Lee J. Cronbach was one of the methodological giants of our field. His contributions include Cronbach’s coefficient alpha, generalizability theory, and notions about construct validity. As his strong roots in methodology and social science research significantly influenced his evaluation approach—evaluation for improving social systems—we position this theory on the methods branch of the theory tree. This evaluation approach has elements that reflect a broader view of evaluation than those presented in experimental approaches. To begin, Cronbach was a student of Ralph Tyler; his orientation to improvement is particularly salient in Cronbach’s approach, unlike other experimental evaluation approaches.

Following the work of Carol Weiss, Cronbach rejected the simplistic model that assumes a single decision maker and “go/no-go” decisions. Instead, he viewed evaluation as an integral part of policy research focused on “policy-shaping communities” and necessitating potential political accommodation. Thus, while there is discussion of

the evaluator’s active role in providing data to local decision makers for instrumental use in accordance with contractual obligations, this theoretical approach is more oriented toward evaluation utilization for enlightenment purposes than toward a concern for immediate instrumental use.

To define the manner in which evaluators might most productively enhance enlightenment use, a set of symbols is used to define four domains of evaluation (Cronbach, 1982). These domains consist of *units* (populations), *treatments*, *observations* (outcomes), and *settings*. The sample examined by the evaluator, referred to by the acronym UTOS, represents the larger population from which sampling took place. Cronbach argued the most important area of concern for evaluation is external validity, which is described as the plausibility that the conclusions to UTOS are different from the population under study, referred to as *UTOS (pronounced “star-UTOS”).

With concern for evaluations contributing to enlightened discussion, evaluation for improving social systems focuses on what Cronbach referred to as “bandwidth,” meaning that it is more important for the evaluation to focus on a broad range of relevant issues than to achieve absolute fidelity (accuracy) on a small number of issues. He explained that generalization to *UTOS can be attained by extrapolating through causal explanation, either using causal modeling or the “thick description” of qualitative methods. Furthermore, it is sometimes beneficial to examine subpopulations (sub-UTOS) because focusing on the subset of data for a particular group might enable generalization to other domains. These arguments are an extension of Cronbach’s earlier research on aptitude–treatment interactions conducted jointly with Richard Snow (Cronbach & Snow, 1977).

A key assumption of this approach is a sensitivity to the values of the policy-shaping community, arguing that the evaluator should seek to incorporate their views on the evaluation questions most worth

addressing. This is done systematically with an eye toward what will contribute most to generalization and to the issues that are receiving attention from the policy-shaping community, that are most relevant in swaying important (or uncommitted) groups, that have the greatest uncertainty, and that would best clarify why a program works. Shadish and colleagues (1991, p. 340) made the following insightful distinction between Cronbach's theoretical approach and several other major approaches: "[Cronbach views] evaluators [as] educators rather than [as] the philosopher-kings of Scriven, the guardians of truth of Campbell or the servants of management of Wholey" (p. 340).

As such, the evaluation for improving social systems approach does not aspire to the instrumental use of theories on the use branch, nor does it argue for the strict experimental controls laid out in Campbell's writings on experimentation. Likewise, the evaluator is not called upon to impose their own value system on the program or to reach summary judgments about programs. Rather, the evaluator's role is to provide "readers" with information that they may consider when forming their own judgments.

We have placed evaluation for improving social systems, as advanced by Cronbach and his associates, on the methods branch of the evaluation tree in a position that is in the direction of the use branch. This placement in part reflects the influence of Carol Weiss's use-focused writings on this approach with respect to enlightenment use directed toward policy-shaping communities.

Tailored Evaluation: Peter Rossi

Peter Rossi is well known for his highly popular textbook on evaluation. His first edition depicted and relied most heavily on design considerations from Campbell and Stanley (1966), as reflected in Suchman's (1967) influential evaluation book, which focused on evaluation as research. Thus, the title of Rossi's first edition was *Evaluation: A Sys-*

tematic Approach (Rossi, Freeman, & Wright, 1979). Now in its eighth edition—with modified coauthors (Rossi, Lipsey, & Henry, 2019)—it is still a very popular and influential text. Over time, however, because of the influences of other evaluation theorists, and as other perspectives and the changing nature of the field have modified views, the text now includes discussions of qualitative data collection, evaluation utilization, the role of stakeholders, and so on.

While Rossi's earlier writings stressed the use of the experimental design, to describe him as exclusively an experimentalist would be erroneous. He contributed significantly to the conceptualizations of theory-driven evaluation and comprehensive evaluation, which he described as "the systematic application of social research procedures in assessing conceptualization and design, implementation, and utility of social intervention programs" (Rossi & Freeman, 1985, p. 19). However, as is evident from this quote, while he moved away from considering the experimental model as the *sine qua non* of evaluation, he still clearly viewed evaluation as social research.

Rossi's departure from the randomized experiment was initiated in part by his broadening vision of evaluation. Concerning himself with issues such as implementation, use, and internal and external validity, his ideas eventually evolved to a place that some suggest was so comprehensive that the approach was virtually impossible to implement. Rossi's response to this criticism led him to develop the approach he referred to as *tailored evaluation*, which assisted in refining the evaluation focus. In this work, Rossi suggested that methods be tailored to the stage of the program—that is, "fitting evaluations to programs" (Rossi & Freeman, 1985, p. 102). For example, if commissioned to conduct an evaluation of a program that was well underway, the evaluator would focus on current work rather than on the development of the program.

Rossi, along with Weiss and Chen, provided some of the foundational think-

ing about theory-driven evaluation. This approach can be seen as a particular type of comprehensive evaluation. Rossi maintained that this approach of detailing a program theory to guide the evaluation helps to reconcile the two main types of validity: internal and external. Since it is difficult to design studies that emphasize both, many theorists believe that one must be traded for the other. Nevertheless, both are important.

Rossi's earlier work focusing on the use of the experimental design was obviously more concerned with internal than with external validity. However, he began to question whether internal validity should have priority. Rossi asserted that this approach was a fusion of Campbell's experimental evaluation and Cronbach's evaluation for improving social systems, as it addressed internal validity through a specified program model and external validity through an investigation of the ways the program was similar to and different from future programs to which the evaluation was to be generalized. Shadish and colleagues (1991) said that comprehensive evaluation presents a multifaceted, broad-based point of view to appeal to a wide audience. We believe it more accurate to say that a recent edition of Rossi's text (Rossi et al., 2019) reflects a concern for demonstrating great breadth and for the depiction of evaluation possibilities to capture the interests of broader audiences.

Evidence-Based Policy Use: Carol Weiss

Carol Weiss made so many important contributions to evaluation and has influenced our thinking about what evaluation can do. Much of her thinking was informed by the research she conducted on evaluation use in the context of political decision making. Through her empirical work, Weiss expanded on, defended, or made precise many of the key concepts and terms related to evaluation use, perhaps most notably *conceptual use*, or use for understanding (Weiss, 1979, 1981); *enlightenment use*, or more subtle and indirect use that occurs in the lon-

ger term (Weiss & Bucuvalas, 1980); and, with the emergence of evaluation policy (see Christie & Lemire, 2019), *imposed use*, a requirement by federal funding agencies that grant recipients only run programs that have reported scientific evidence of success (Weiss, Murphy-Graham, & Birkeland, 2005; Weiss, Murphy-Graham, Petrosini, & Gandhi, 2008). Based on her empirical work, Weiss argued for more evidence-based policymaking and concluded that the most effective kinds of evaluations are those that withstand the test of time—that is, are generalizable and therefore use the most rigorous methods possible. Thus, Weiss's empirical work, which examines evaluation use, informed her prescriptive theoretical work, where she argued for the use of research methods to inform evidence-based policy decisions.

Weiss's early evaluation work was influenced by research methodologists, political theorists, and expositors of democratic thought (e.g., Rousseau, the *Federalist Papers*). Weiss's firm commitment to evaluation as being akin to carefully conducted research is documented in many of her writings. In fact, her first evaluation text, *Evaluation Research: Methods of Assessing Program Effectiveness* (1972b), was intended to be a supplementary text for research methods courses. Shadish and colleagues (1991) pointed out that Weiss's early writings focused almost exclusively on "traditional experimental methods in evaluation" (p. 183). She later argued a similar position: "Evaluation is a kind of policy study, and the boundaries are very blurred. . . . I think we have a responsibility to do very sound, thorough systematic inquiries" (Weiss, cited in Alkin, 1990, p. 90).

Throughout her writings, Weiss demonstrated an appreciation for evaluation as a political activity. Her notion of evaluation extended to how study findings affect and are affected by the political context. Weiss (1991, p. 213) was particularly pointed in recognizing that the process of conducting an evaluation affects a political situation in

which there are vested interests, negotiation, constituents, critics, and so on, saying, “Politics intrudes on program evaluation in three ways: (1) programs are created and maintained by political forces; (2) higher echelons of government, which make decisions about programs, are embedded in politics; and (3) the very act of evaluation has political connotations.”

Weiss recognized the complexity of this political context. Thus, she was adamant in noting that evaluation results are not necessarily (or generally) heeded in isolation from other input or at the time of evaluation reporting. She is known for, among many other contributions, introducing the term *decision accretion*, meaning that decisions are the result of “the build-up of small choices, the closing of small options and the gradual narrowing of available alternatives” (Weiss, 1976, p. 226).

Given the belief that decisions accrete, Weiss studied policy research to determine the characteristics of evaluative research work that have longevity—that is, those that would be considered credible during an accretion process. She concluded that research quality is of greatest importance “not only because of adherents to the norms of science, but perhaps more importantly because it increases the power of research as ammunition in intra-organizational argument” (Weiss & Bucuvalas, 1980, p. 256). It is this type of “careful and sound research” (Weiss, 1981, p. 400) that she continued to argue is the most important kind of evidence an evaluation can offer policymakers.

Weiss’s views on evaluation place limits on the likelihood of instrumental use; instead, she viewed evaluation’s greatest contribution as conceptual or enlightenment. Thus, evidence-based evaluation use does not fit within the use branch but rather is placed on the methods branch. As Shadish and colleagues (1991, p. 208) noted about Weiss, “When forced to choose between traditional scientific rigor versus other goals such as use, she seems to lean in the former direction.”

Valuing

Out of the root of epistemology has grown a branch of evaluation approaches that focuses on concerns related to valuing in the evaluation process. Of particular importance is the fact–value distinction delineated by 18th-century Scottish philosopher David Hume. To some extent, the issue of what is “fact” has been addressed and amplified using a strict distinction between fact and value in the second root, social inquiry. The variance from this view of “fact” is important itself as a root. Important issues raised when considering valuing in evaluation include the legitimacy of value claims, as ably described by House and Howe (1999); the nature of universal (justifiable) claims; and the constructivist view that truth (or fact) is guided by “the meanings that people construct in particular times and places” (Greene, 2009, p. 159).

To better reflect the philosophical theories informing the evaluation approaches on this branch, we have split it in two, naming the left arm stretching toward the methods branch “objectivist influenced” and the right arm “subjectivist influenced.” Again, we underscore the use of the word *influence* here when describing the valuing branch. At the base, prior to the split of the branch, are goal-free evaluation (Scriven, 1967, 1972a) and responsive evaluation (Stake, 1967, 1974). These two theories have significantly influenced the field in terms of both its conceptual underpinnings as well as its procedures for practice.

This branch also includes the work of those who reject the notion that we should strive for an objectivist judgment about the merit or worth of the evaluand. Their approaches espouse the philosophy of relativism or subjectivism—that is, the claim that human activity is not like that in the physical world, but rather, reality is an ongoing, dynamic process. This is quite different from the type of valuing that concerns Scriven. Stake’s 1967 paper, “The Countenance of Educational Evaluation,”

offers hints of subjectivist thinking, but he explicitly rejected “preordinate evaluation” in his 1974 paper, “Program Evaluation, Particularly Responsive Evaluation” (evaluation conducted in the spirit of obtaining objective information). There, he argued for a responsive approach using case studies as a means for (1) capturing the issues, personal relationships, and the complexity of the evaluand and (2) judging the value of the evaluand under study. Stake’s early subjectivist conceptualization of an approach to evaluation has served as an important influence on the thinking of others who argue for attention to complexity, dialogue, and meaning in evaluations and use this as a basis for informing value claims in evaluation studies.

Goal-Free Evaluation: Michael Scriven

Michael Scriven (1972b) advocates for *goal-free evaluation*, in which the evaluator assumes responsibility for determining which program outcomes to examine, rejecting the objectives of the program as a starting point. He has maintained that, by doing so, the evaluator is better able to identify the real accomplishments (and nonaccomplishments) of the program. An essential element of this valuing approach is the determination of a single value judgment of the program’s worth (“good” or “bad”). In requiring a synthesis of multiple-outcome judgments into a single value statement, Scriven is alone among evaluation theorists.

In both goal-free evaluation and the synthesis stage, Scriven has justified his point of view by relying on the extent to which the program is able to meet “needs,” the presumed cost to society and individuals as determined by a needs assessment. Shadish and colleagues (1991) maintained that Scriven’s conception of valuing is dependent on his definition of needs and noted that his “conception of needs implies a prescriptive theory of valuing and that he disparages descriptive statements about what people think about the program” (p. 95). Moreover,

they maintained that his needs assessment is not independent of the views of the evaluator and that failing to directly reflect the views of stakeholders inhibits the potential use of evaluation findings in policymaking. Scriven is apparently unconcerned by this, maintaining that determining “truth” is sufficient. Scriven’s unique training in philosophy, mathematics, and mathematical logic provides him with the assurance that he can make sound, unbiased judgments.

Scriven’s major contribution is the way in which he has adamantly defined the role of the evaluator in making value judgments. Scriven has been unequivocal in his position that society requires valuing and that evaluation is the science of valuing: “Bad is bad and good is good and it is the job of evaluators to decide which is which” (1986, p. 19) rather than “pass the buck [for final judgment] to the non-professional” (Scriven, 1983, p. 248).

The evaluator, in valuing, must fulfill his or her role in serving the public interest (Scriven, 1975). By public interest, Scriven does not simply refer to clients, users, or stakeholders, but to all potential consumers. Indeed, he views the evaluator’s role in valuing as similar to producing a study for *Consumer Reports* in which the evaluator determines the appropriate criteria by which judgments are to be made, then presents these judgments for all to see. As in *Consumer Reports*, evaluators must identify “critical competitors,” or competing alternatives. Just as *Consumer Reports* would identify “midsize cars” to be evaluated, so, too, would the evaluator seek out similar entities for evaluation.

Scriven has adamantly stated that it is not necessary to explain why a program or product works in order to determine its value. Nevertheless, he introduced an alternative to experimental and quasi-experimental design called the “modus operandi” (MO; Scriven, 1991, p. 234), which is analogous to procedures used to profile criminal behavior: “The MO of a particular cause is an associated configuration of events, processes, or

properties, usually in time sequence, which can often be described as the *characteristic causal chain* (or certain distinctive features of this chain) connecting the cause with the effect” (Scriven, 1974, p. 71).

In a manner similar to that which Scriven has used to grade critical competitors, the MO method requires that the evaluator first develop a thorough list of potential causes and then narrow it down by determining which potential causes were present prior to the effect and then which complete MO fits the chain of events, thus determining the true cause. To ensure accuracy and bias control, the evaluator looks for instances of “co-causation and over determination,” and calls in a “goal-free or social process expert consultant to seek undesirable effects” (Scriven, 1974, p. 76). Scriven believes that, ultimately, the evaluator delivers a picture of the causal connections and effects that eliminate causal competitors without introducing evaluator bias.

Extending his supposition for evaluation as the science of valuing, Scriven (1991, 2001, 2003) has reasoned that evaluation is a *transdiscipline*—that is, a discipline that possesses its own unique knowledge base while serving other disciplines as a tool. He has maintained that, like logic and statistics, evaluation is a major transdiscipline because all disciplines rely on the evaluation process to judge the value of the entities within their own purview. This view of evaluation gave rise to Scriven describing evaluation as an “alpha discipline.”

Scriven’s thinking pushed the field to consider valuing as a central feature of evaluation more so than anyone else. Thus, his approach is positioned on the valuing branch at the base of the objectivist influence arm.

Educational Connoisseurship: Elliott Eisner

Elliott Eisner first presented his views on what he called *educational connoisseurship* in an issue of the *Journal of Aesthetic Education* (1976), and he subsequently expanded on

those views (1985, 1991a, 1991b, 1998) in several important ways. His evaluative viewpoint was a direct response to his criticism of more traditional approaches to evaluation focused on educational outcomes, whether measured by standardized tests using principles of psychological testing or by criterion-referenced testing procedures. Eisner’s (1976) rejection of “technological scientism” included a rejection of the extensive use of research models employing experimental and quasi-experimental designs, which are heavily (if not exclusively) dependent on quantitative methods, noting that “things that matter” cannot be measured quantitatively. Eisner argued that while quantitative techniques can provide some useful information, “evaluation requires a sophisticated, interpretive map not only to separate what is trivial from what is significant, but also to understand the meaning of what is known” (1994, p. 193).

Drawing heavily from his prior experience as both a curriculum expert and artist, Eisner used the role of critics in the arts as an analogy for an alternative conception of evaluation. Central to his evaluation theoretic views are the twin notions of connoisseurship and criticism. To be a connoisseur is to have knowledge about what one sees, to have the ability to differentiate subtleties, and to be aware of and understand the experience. Eisner (1991b, p. 174) noted that “a connoisseur is someone who has worked at the business of learning how to see, to hear, to read the image or text and who, as a result, can experience more of the work’s qualities than most of us.”

Criticism, in Eisner’s view, is making the experience public through some form of representation. First, there is *critical description*, in which the evaluator draws on [their] senses to describe events, reactions, interactions, and everything else that is seen. By doing this, the evaluator portrays a picture of the program situation, frequently imagining themselves as a participant and drawing on the senses to describe the feeling in the participants’ terms. The second dimension

of criticism is *expectation*, in which the evaluator tries to understand or make sense of what was seen. The evaluator asks, "What is the meaning of this?" and "What message are participants getting?" In responding to these questions, the evaluator must draw on their own experience to understand the minds of participants. Eisner emphasized the necessity of background knowledge since a good critic cannot interpret or evaluate what they have not experienced themselves. The critic must ask, "What is the value of what is happening?" Eisner (1991b) noted that valuing is a critical element in the process:

The essence of perception is its selectivity; the connoisseur is as unlikely to describe everything in sight as a gourmet chef is to use everything in his pantry. The selective process is influenced by the value one brings to the classroom. What the observer cares about, she is likely to look for. . . . Making value judgments about the educational import of what has been seen and rendered is one of the critical features of educational criticism. (p. 176)

Like Scriven, Eisner considers the valuing role important and has expanded it to include not only making final judgments about data (or observations) but also making judgments about what questions to ask and where to focus the evaluation study. Eisner and Scriven have both posited the evaluator as the expert who thus determines the final value of a program. Eisner's approach differs from Scriven's, though, in that he believes the evaluator has the authority to judge a program's merit because of expert understanding of the subject area (e.g., education or public health), connoisseurship, and experience or "the enlightened eye" (1991a) rather than because of expertise as an evaluator.

When on the tree, educational connoisseurship is positioned toward the objectivist side of the valuing branch because of the use of intense observation, which includes both quantitative and qualitative measures but emphasizes the qualitative.

Cost Analyses: Henry Levin

Cost analyses are a critical domain of evaluation work because they offer information to inform what some consider to be the ultimate evaluation question: What is the overall value of the program? Cost analyses are economics-based strategies for determining the value of a program or policy. Henry Levin is a leading and noteworthy educational policy researcher and evaluator who for several decades has focused on developing, promoting, and using cost analyses as a means for drawing evaluative conclusions about programs.

Levin has offered evaluators an array of economics-based strategies for determining costs prior to and during program implementation (Levin, 2005; Levin & McEwan, 2001). In a program's planning stages, cost analysis strategies can be used to project what a program might cost and whether it is fiscally responsible to implement. Likewise, a cost analysis procedure might be used to keep track of the costs of an ongoing program. Cost-effectiveness procedures determine which program out of many achieves a target outcome most frugally or which program of many with equal costs produces the greatest outcome. Furthermore, a program's cost may be examined relative to its monetary impact (or benefits) on clients or society (cost-benefit analysis). An essential part of preparing to undertake a cost study involves thinking about how "costs" are defined, what costs are important, and how best to assess them. Because costs can be calculated and evaluated differently across strategies, it is important for evaluators and stakeholders to clarify what they need from a cost analysis strategy and to understand how each works.

The cost analysis approach is positioned on the valuing branch in the direction of the methods branch. This reflects the primary object of cost analyses as a method for informing object value judgments about a program using a specific methodological approach.

Responsive Evaluation: Robert Stake

Early in the evaluation field's development, Robert Stake authored three manuscripts—*The Countenance of Educational Evaluation* (1967), *Program Evaluation, Particularly Responsive Evaluation* (1974), and *Case Studies in Science Education* (with Easley & Anastasiou, 1979)—that served as the primary foundation for his later work. These works served as the foundation for his conceptualization of responsive evaluation, which is grounded in three essential beliefs: (1) There is no true value to anything (i.e., knowledge is context-bound); (2) stakeholder perspectives are integral elements in evaluations; and (3) case studies are the best method for representing the beliefs and values of stakeholders and of reporting evaluation results.

Stake describes the case study as an essential method in responsive evaluation, as it is necessary for providing “thick description” of the evaluand. Stake (2001) argued that the judgmental act is part of the descriptive, observational act. In other words, he maintained that seeing and judging the evaluand regularly are part of the same act and that the task of evaluation is as much a matter of refining early perceptions of quality as of building a body of evidence to determine the level of quality.

In taking a responsive evaluation approach, Stake argued against stakeholder participation in many evaluation processes and instead asserted that evaluation is the job of the evaluator (Alkin, Hofstetter, & Ai, 1998, p. 98). Stake (2000) noted that it is the evaluator's job “to hear the [participants'] pleas, to deliberate, sometimes to negotiate, but regularly, non-democratically, to decide what [the participants'] interests are” (p. 104). Stake (1975) cautioned that “whatever consensus in values there is [among participants] . . . should be discovered; the evaluator should not create a consensus that does not exist” (pp. 25–26). Furthermore, he noted that “the reader, the client, the people outside need to be in a position to

make their own judgments using grounds they have already, plus the new data” (Abma & Stake, 2001, p. 10).

Stake, like Campbell, Cronbach, and Scriven, is one of the evaluation field's pioneers and his work has influenced all aspects of evaluation theory. Stake's attention to the notion of stakeholder participation in responsive evaluation and the visibility that he accorded these ideas has had a strong influence on theories that we have positioned on the use branch of the evaluation theory tree. Responsive evaluation differs from use theories, however, in that the motivation for including stakeholders is not centered on an effort to promote use, nor is there particular attention given to decision makers or to how evaluation findings might be used.

Placing responsive evaluation on the valuing branch provides an easy transition to House's deliberative democratic evaluation approach, which embraces the role of the stakeholder in an evaluation but focuses explicitly on giving voice to those who are underrepresented in the evaluation process. The subjectivist arguments underpinning responsive evaluation are also evident in the work of Guba and Lincoln (1989), whose fourth-generation evaluation approach is centered in constructivist philosophy. Thus, we place responsive evaluation on the valuing branch, acknowledging its substantial influence on theories across the tree.

Deliberative Democratic Evaluation: Ernest House

Ernest House recognized that evaluation serves the purpose of providing information to decision makers so that they can determine the legitimate allocation of vital resources. He denounced as inadequate the utilitarian framework that underlies key theories of evaluation (primarily theories on the methods branch and early theories on the use branch). House (1991, p. 235) noted that “utilitarianism is a moral theory which holds that policies are morally right when

they promote the greatest sum total of good or happiness from among the alternatives.” In the pursuit of social justice, he argued that attempts at being responsive to stakeholders are superior to prior evaluation conceptions. He also criticized the lack of value neutrality in stakeholder approaches, which he said results from the general lack of full inclusion of the represented interests of the poor and powerless in stakeholder groups. House argued that it is the evaluator’s role to prioritize the voices of stakeholders least likely to be heard in an evaluation rather than those of key decision makers.

House came to these views by drawing on Rawls’s (1971) justice theory. House (1991, 1993) argued that evaluation is never value-neutral and should tilt in the direction of social justice by specifically addressing the needs and interests of the powerless. House’s evaluator is thus faced with the task of understanding the needs and positions of various stakeholder groups, especially the powerless, and of balancing this information with their own perceptions of social justice. In doing this balancing, the evaluator shapes the kind of information collected, as well as its analysis. In essence, evaluators cast themselves in the role of spokespeople or representatives of the poor and powerless. For House, the role of the evaluator is not to define value in terms of good or bad, as Scriven has argued, but in terms of truth, beauty, and justice. In this sense, the value judgments according to the least advantaged should receive the utmost importance.

It is informative to examine what House has referred to as “ethical fallacies” within evaluation: *clientism* (taking the client’s interest as the ultimate consideration); *contractualism* (adhering inflexibly to the contract); *managerialism* (placing the interest of the managers above all else); *methodologicalism* (believing that proper methodology solves all ethical problems); *pluralism/elitism* (including only the powerful stakeholders’ interests in the evaluation); and *relativism* (taking all viewpoints as having equal merit). By referring to these as ethical fal-

lacies, House disassociated himself from the theoretic goals that are consistent with each. For example, the first three can easily be associated with the use branch of the theory tree (particularly the early part of the branch); methodologicalism is clearly associated with the methods branch; and pluralism/elitism and relativism are associated with various stakeholder approaches.

A bit later in his career, House began writing with Kenneth Howe; it is in this writing that we find some of his most prescriptive thinking on how evaluation should be conducted. House believed that “evaluators should accept authority but not power” (House & Howe, 1999, p. 102). An evaluator who caters to those with power perpetuates inequality, which is why inclusion is the first criterion of the deliberative democratic evaluation approach. Inclusion addresses stakeholder bias (of those in power) and invites stakeholders with and without power to participate in the evaluation. Dialogue, the next criterion, is necessary to ensure that stakeholder contributions are well thought out and can be honestly weighed along with all other stakeholder contributions to the evaluation. Deliberation, the final criterion, is the weighing of each contribution to generate an accurate conclusion.

House and Howe’s approach is value-engaged. Fact and value are not mutually exclusive; they exist on a continuum where a middle ground exists between “brute fact” and “bare values” (House & Howe, 1999, p. 6). The deliberative democratic evaluation process is described as follows: “We can imagine moving along the value–fact continuum from statements of preferences and values collected through initial dialogue, through deliberations based on democratic principles, to evaluative statements of fact” (House & Howe, 1999, p. 100).

House is often cited as the theorist who first acknowledged the ways in which evaluation impacts power and social structures and described how it can be used to either shift or maintain existing repressive structures. His work reminds us that evaluation

is used to determine “who gets what” and that, without explicit concern for those from underrepresented groups, evaluation is not meeting its primary purpose—to promote social justice. House’s emphatic concern with values and judgment in his theoretical approach firmly positions deliberative democratic evaluation on the valuing branch of the evaluation tree.

*Values-Engaged Evaluation:
Jennifer C. Greene*

Jennifer Greene has argued that evaluation should be used to determine value, specifically by developing a consensus around a set of criteria used to determine the value of a program. Greene (2005) described what she calls a value-engaged approach to evaluation, which incorporates elements of responsive evaluation (Stake, 1975, 2005) along with principles from deliberative democratic evaluation (House & Howe, 1999), including its three criteria: inclusion, dialogue, and deliberation. Value-engaged evaluation also stresses stakeholder involvement, which closely resembles participatory evaluation approaches, and the use of mixed-methods designs and fieldwork in evaluation.

Value-engaged evaluation emphasizes responsiveness to the particularities of the context, inclusion of and engagement with multiple stakeholder perspectives and experiences, and attention to the social and relational dimensions of evaluation practice (Greene, 2005). This approach is responsive in that the evaluation design is developed in response to the program context. It is democratic in that it places great importance on considering the perspectives and values of all legitimate stakeholders, especially those who are typically alienated from evaluation processes. Greene (2000) argued that evaluations can be “used to surface and legitimate differing views and values and move [stakeholders] towards shared understanding of the values of educational outcomes” (p. 16).

Greene offered three “justifications” for including stakeholder views when conducting evaluations: pragmatic, emancipatory, and deliberative. The *pragmatic* justification argues for stakeholder inclusion because it increases the chance of evaluation utilization and organizational learning. The *emancipatory* justification focuses on the importance of acknowledging the skills and contributions of stakeholders and empowering them to be their own social change agents. The *deliberative* justification argues that evaluation should serve to ensure that program or policy conversations include all relevant interests and are “based on the democratic principles of fairness and equity and on democratic discourse that is dialogic and deliberative” (Greene, 2000, p. 14). Implementation of these principles is intended to support impartiality through the inclusion of pertinent and germane interests, values, and views to reduce biases in values and facts that are used to inform study conclusions (House & Howe, 2000).

Greene has been explicit about the extent to which Stake and House have influenced her work. The impact of their work on her own ideas is evident throughout her writings. Placing the values-engaged evaluation approach on the valuing branch after deliberative democratic evaluation offers a visual depiction of the relationship between the approaches and the influence of key theory writers on Greene’s development of this model.

*Culturally Responsive Evaluation:
Rodney Hopson*

Culturally responsive evaluation (CRE) is centered on incorporating program cultural context into the evaluation design and process (Hood, Hopson, & Kirkhart, 2015). Rodney Hopson and Stafford Hood were early writers on this approach and have contributed significantly to the development of the principles and practices of CRE. In this approach, conceptualizations of culture go beyond the demographic characteristics

of the program community. Rather, evaluators should be aware of larger societal structures that affect those involved in the study, including systems of race, class, and gender (SenGupta, Hopson, & Thompson-Robinson, 2004).

Often, evaluators work in cross-cultural contexts where the funder, evaluator, and program stakeholders are of different cultural backgrounds. Instead of imposing a set of cultural views onto a community, in a culturally responsive approach, the evaluator should seek to understand and incorporate the program's cultural context into the evaluation. Along the same lines, a culturally competent evaluator must examine how their own positionality, power, and privilege may influence the evaluation practice (American Evaluation Association, 2011). Embedded in this approach is the acknowledgment that culture shapes values and beliefs. Therefore, to authentically represent the value, merit, and worth of a program, the evaluation must be responsive to the program's cultural context (SenGupta et al., 2004).

To incorporate a program's cultural context, culturally responsive evaluators involve program stakeholders throughout the evaluation process, including in crafting evaluation questions, developing a methodological approach, and determining measures of success. When deciding which stakeholders to involve, evaluators should seek those whose voices may be less heard, such as underrepresented and vulnerable populations (Hood et al., 2015). The result of active stakeholder involvement is an evaluation that is both epistemologically and methodologically culturally congruent. This implies flexibility in how the evaluation is designed to be reflective of the community's values. Therefore, epistemologically, culturally attuned methods of evaluation reject the conception of a single, overarching truth and instead stress that knowledge must be understood within the historical, social, and cultural context of the community being studied. As a result, a culturally responsive approach can be

used to respond to unequal power relations within the evaluation process that may marginalize or exclude the knowledge of the community participating in the evaluation (Tillman, 2002).

The integration of culture can lead to studies that are aligned with community values and responsive to the goals and needs of those affected by the evaluation. Doing so can increase the validity and usefulness of findings to the community (Kirkhart, 2010; Trickett & Espino, 2004). By prioritizing the values and experiences of the program community, culturally responsive approaches aim to use evaluation as a tool to promote equity (American Evaluation Association, 2011).

Transformative Evaluation: Donna Mertens

Inclusive/transformational evaluation (Mertens, 1999, 2009) has strong grounding in the principles of fourth-generation evaluation and deliberative democratic evaluation, but it is placed within the group of theories that identify cultural relevance as a primary emphasis of evaluation. The theory has a strong emphasis on diversity, the inclusion of diverse groups, and the promotion of equity through intentional work against oppressive systems. The inclusive/transformational evaluation approach emphasizes the transformational paradigm, which comprises four philosophical assumptions: the nature of reality, the nature of knowledge and the relationship between the researcher and the community, the nature of inquiry, and the nature of ethics. The primary focus is on the fourth assumption—the nature of ethics—which specifically relates to social justice and human rights (Mertens & Hesse-Biber, 2013).

In this model, the evaluator's primary role is to include marginalized groups rather than to act as the decision maker (Mertens, 2009). The evaluator makes an intentional effort to be culturally responsive, address inequities, bring forth knowledge surrounding oppression, and design methods that

best fit the needs of all stakeholder groups, with special attention to power differences. The evaluator works with stakeholders to ensure that no decisions are made regarding problems and solutions without their consultation. The evaluator seeks to establish relationships of trust with stakeholders while fostering evaluation capacity building.

Mertens has asserted that evaluators working within an inclusive framework should ask themselves the following questions at the planning stages of the evaluation:

- Are we including people with diverse abilities, ages, classes, cultures, ethnicities, gender identities, families, incomes, languages, locations, races, and sexualities? What barriers are we erecting to exclude a diversity of people?
- How may preexisting oppressive systems play a role in the evaluation? How can we counteract them to address power differences and value grassroots knowledge?
- Have we chosen the appropriate data collection strategies for diverse groups, including providing for preferred modes of communication?

The aim of inclusion is to ensure that the evaluation is conducted within the context of the program's entirety in a way that discourages bias, not to advocate for one group over another. A primary goal of inclusive/transformational evaluation is to challenge the status quo in a quest to transform society through both local and global evaluations. Mertens agrees with and has quoted Chelimsky, who argued that challenging the status quo is "our most important task and the best justification for our work" (quoted in Mertens, 1999, p. 2). In this way, inclusive/transformational evaluations are a tool to confront social inequity and oppressive paradigms and promote equality. Thus, inclusive/transformational evaluation is placed on the valuing branch of the tree, as it explicitly addresses the evaluator's motivation toward and role in the pursuit of social justice.

Indigenous Evaluation: Fiona Cram and Jill Chouinard

Indigenous evaluation is an approach that emerged from culturally responsive evaluation to address the specific needs of Indigenous communities. Inherent in this approach is the recognition of Indigenous peoples' experiences with colonization, their political status as sovereign nations, and their right to self-determination. Its development is in response to a history of research and evaluation work that did not benefit Indigenous communities (Bowman-Waapalaneexkweew, 2018). Research that used a deficit lens, imposed culturally incongruent frameworks, and employed unethical practices led to the misrepresentation and exploitation of Indigenous people (LaFrance & Nichols, 2010; Smith, 2012). In contrast, an Indigenous evaluation approach centers on developing strategies that promote community involvement, cultural vitality, and building evaluation capacity (Chouinard & Cram, 2020).

Building strong relationships with community stakeholders is a cornerstone of Indigenous evaluation. In this approach, the evaluator works closely with community members to incorporate their knowledge throughout the evaluation process and to build credibility and trustworthiness (Kovach, 2010). Likewise, given the history of intrusive research conducted on Indigenous peoples, evaluators have a responsibility to be "relationally accountable" to the community involved and to develop an ethically sound evaluation (Kovach, 2010). This means that evaluators should seek stakeholder insights during each step in the process (Bowman-Waapalaneexkweew, 2018). In particular, evaluators should consult with community members before the evaluation begins to determine their goals, expectations, and willingness to participate and involve stakeholders throughout the process (Bowman-Waapalaneexkweew, 2018; LaFrance, Nichols, & Kirkhart, 2012). Evaluations that do not include authentic

stakeholder involvement or those that do not seek to give back to the community in a meaningful way can perpetuate colonizing research practices. Additionally, stakeholder involvement can be used to build community evaluation capacity, thereby promoting a self-determination agenda (Robertson, Jorgensen, & Garrow, 2004).

Participatory and empowerment forms of evaluation are often aligned with Indigenous approaches. Central to these methods is the community's role in determining the direction of the evaluation and developing their knowledge and skills to continue evaluation efforts after the contract ends. The use of an Indigenous evaluation approach is essential to developing evaluations that challenge colonization and support self-determination efforts.

Fourth-Generation Evaluation: Egon Guba and Yvonna Lincoln

Egon Guba and Yvonna Lincoln view stakeholders as the primary individuals involved in placing value, unlike others on the valuing branch, who view the evaluator as the "valuer." This viewpoint rests on the belief that, instead of just one reality, there are multiple realities based on the perceptions and interpretations of individuals involved in the program to be evaluated. Thus, Guba and Lincoln believe that the role of the evaluator is to facilitate negotiations between individuals reflecting these multiple realities.

Fourth-generation evaluation (Guba & Lincoln, 1989) is based on a constructivist paradigm: That is, in place of a single reality, individuals "construct" their perceptions of reality. The role of the constructivist investigator is to tease out these constructions and "to bring them into conjunction . . . with one another and with whatever information . . . can be brought to bear on the issues involved" (p. 142). More specifically, evaluators are "orchestrators of a negotiation process that aims to culminate in consensus on better informed and more sophisticated constructions" (p. 110).

Guba and Lincoln (1989) described fourth-generation evaluation as "a marriage of responsive focusing—using the claims, concerns, and issues of stakeholders as the organizing elements—and constructivist methodology aiming to develop judgmental consensus among stakeholders who earlier held different, perhaps conflicting, emic constructions" (p. 184). Thus, they argued for using maximum variation sampling to identify the broadest scope of stakeholders. These stakeholders are then interviewed sequentially in order to place on the table the great variety of individual constructions. This is part of what they referred to as "the hermeneutic circle," which defines a continuous interplay of data collection and analysis. Ideally, out of this process, a joint construction will begin to emerge. Additional roles for the evaluator include "testing and enlarging within-group constructions by introducing new or additional information"; "sorting out resolved CC&I [claims, concerns and issues]"; "prioritizing unresolved CC&I"; "collecting information bearing on unresolved CC&I"; "preparing an agenda for negotiation"; "carrying out a negotiation"; and "reporting via the case study—the joint construction as product" (Guba & Lincoln, 1989, p. 185).

Aside from emphasis on bringing together constructions to facilitate valuing by stakeholders, there are numerous places in the process where the evaluator engages in valuing. Indeed, according to these theorists, "'facts' and 'values' are interdependent. 'Facts' have no meaning except within some value framework; they are value-laden. There can be no separate observational and valuational languages" (Guba & Lincoln, 1989, p. 105). A number of Guba and Lincoln's comments seem to validate the point that fourth-generation evaluators do play at least a moderate personal role in valuing. Note the following:

As the several constructions begin to take shape, however, certain elements will seem to be more salient than others (and will probably first appear this way to the inquirer). (p. 153)

The human is the instrument of choice for the constructivist for only a human can enter a context without prior programming, [and,] after a short period, begin to discern what *is* salient (in the emic view of the respondents). (p. 175)

It is clear that fourth-generation evaluation fits well on the valuing branch of the evaluation theory tree. It is particularly interesting to observe among the various approaches on this branch the great diversity in the manner in which valuing takes place. Yet all share a primary concern for valuing, either by the evaluator or through the evaluator acting in a role designed to facilitate valuing by others.

Equitable Evaluation: Asa Hilliard III

The contributions of African American and Black scholars to evaluation have not been properly acknowledged, documented, and recognized. Among these scholars are Aaron Brown, who conducted major evaluation studies; Leander Boykin (1949), who provided a set of guiding principles for effective evaluation; and Asa Hilliard III, whose work shaped our theoretical notions and understandings of equitable evaluation.

Hilliard was primarily a scholar of education, widely influenced by his observations of schools post *Brown vs. Board of Education*. In his work, he incorporated liberation and equity into the education system by shifting the Eurocentric lens through which the history of the African diaspora and of African Americans was taught (Jamison, 2019). This led to Hilliard's analysis of assessment of educational practice and, consequently, his focus on evaluation.

While Hilliard did not specifically reference his theoretical contributions by title or name, here we refer to his work as "equitable evaluation." Hilliard describes evaluation as "being part of a larger process of envisioning social and educational [processes] of democratic participation" (Hood & Hopson, 2008, p. 416). His evaluation theory and approaches to evaluation evolved

in response to his evaluation practice in educational settings. In Hilliard's evaluation writings, he addresses the importance of valuing on contexts that serve predominantly Black communities and other minoritized groups, although his principles may be viewed as widely applicable to many contexts. Hilliard's concern with the complexity of the well-documented biases of large-scale standardized testing would focus his evaluation theory work as mostly applicable to state and local program evaluations.

Hilliard's work provided the foundation for advancing theories such as culturally responsive evaluation and for influencing noted scholars such as Hopson and Hood. In his writings, he argues for the importance of using evaluation as a process for empowering peoples traditionally marginalized in education-related evaluative work and in society more generally. Hilliard prioritized *equity* and democracy in evaluative inquiry, which fostered the subsequent development of culturally responsive evaluation (Hood, Hopson, & Kirkhart, 2015).

We have placed equitable evaluation on the valuing branch of the theory tree as it has grown from notions of responsive evaluation and is closely tied with several other forms of evaluation that center justice in the evaluation process. His theory is oriented toward the use branch, as he sought to address the inequities present in evaluation through practical application in schools for the purpose of improving schooling for Black and other minoritized students and educators.

Use

The use branch began its growth with what are often referred to as *decision-oriented theories*. Decision-oriented theories maintain the core principle that evaluations should be designed specifically to assist key program stakeholders in program decision making. Such stakeholders are most often those that commission the evaluation. Stufflebeam's CIPP model (described below) is one of the best known of these theories. Subsequently, based on empirical knowledge about the

conditions under which evaluation use takes place, utilization theorists have built on the notions advanced in earlier decision-oriented theories. The resulting class of theories is concerned with designing evaluations that are intended not only to inform decision making but also to ensure that evaluation results impact decision making, organizational change, and conceptual understandings of programs. This concern for the use of an evaluation focuses primarily on the program at hand—*this* program at *this* time.

With recent extensions of the use concept in the literature to include evaluation influence, it is important to clarify what we mean by *use* as it relates to this branch of the theory tree. *Evaluation influence* refers to the capacity of evaluation processes, products, or findings to indirectly produce a change in understanding or knowledge, either at the evaluation site at a future time or at other sites (Alkin & Taut, 2003; Christie, 2007; Kirkhart, 2000; Mark & Henry, 2004). Rather than drawing from this broad definition of *use*, the use branch as we envision it depicts the work of theorists primarily concerned with direct program site use (in action or understanding) that results from a particular evaluation study. Accordingly, the theories presented on the use branch are articulated to promote the kind of actionable use that is within the purview of the evaluator.

Context, Input, Process, and Product:
Daniel Stufflebeam

Daniel Stufflebeam, along with Guba, initially developed the CIPP model as an approach to evaluation focused on the decision-making process. CIPP is an acronym for four types of evaluation: context, input, process, and product. *Context evaluation* identifies needs to decide on program objectives. *Input evaluation* leads to decisions about strategies and designs. *Process evaluation* identifies shortcomings in a current program to refine implementation, and *product*

evaluation measures outcomes for decisions regarding the continuation or refocus of the program.

The CIPP model suggests a cyclical process for evaluation wherein the key strategy is to work with a carefully designed evaluation while maintaining flexibility. According to Stufflebeam (1983), evaluators must view design as a process, not a product. Evaluations should provide a continual information stream to decision makers to ensure that programs continually improve their services. To improve services, evaluations should aid decision makers in allocating resources to programs that best serve clients.

CIPP evaluations (Stufflebeam, 2000) are often keyed to *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994), which describes the professional standards by which evaluators should conduct their work and points to four domains of practice: utility, feasibility, propriety, and accuracy. *Utility* standards ensure that an evaluation will serve the information needs of intended users; *feasibility* standards ensure that an evaluation will be realistic, prudent, diplomatic, and frugal; *propriety* standards ensure that an evaluation will be conducted legally, ethically, and with due respect for the welfare of both those involved in the evaluation and those affected by its results; and *accuracy* standards ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. Moreover, the feasibility standard of formal contracts is a way for evaluators and clients to reach formal written agreements that detail “what is to be done, how, by whom, and when so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it” (Stufflebeam, 2000, p. 311).

Using the CIPP approach, an evaluator engages a representative stakeholder panel to help define the evaluation questions, shape evaluation plans, review draft reports,

and disseminate the findings (Stufflebeam, 2001). The stakeholder panel is the primary group with whom the evaluator interfaces regularly; the success of the evaluation can hinge on these regular interactions because it is believed that without them, the evaluation approach will fail. The evaluator keeps the panel abreast of the formative information produced so that decisions about both the program and the evaluation can be made. Evaluations produce a comprehensive assessment of merit and program worth.

In summary, the CIPP model engages stakeholders (usually in decision-making positions) in focusing the evaluation and in making sure that it addresses their most important questions; provides timely, relevant information to assist decision making; and produces an accountability record. Both formative and summative information becomes available to a panel of stakeholders who promote the use of the evaluation findings. Including multiple stakeholder perspectives increases the possibility that relevant value perspectives are represented, thus fostering a comprehensive evaluation of program value. With this as the primary focus, the CIPP model is positioned as the first approach on the use branch of the theory tree.

Results-Oriented Evaluation: Joseph Wholey

From Joseph Wholey's academic training and exposure to large-scale program evaluations, one might anticipate a theory developed by him to be orientated to and aligned with theories classified on the methods branch. However, Wholey's long-standing participation in federal government programs has made him sensitive to the needs of program managers; in many respects, his focus on managers and policymakers (he is less concerned about other stakeholders) is similar to the ways in which the CIPP model orients evaluators to attend to decision makers. Most of Wholey's writings (e.g., Wholey, 1983; Wholey, Hatry, & Newcomer, 2010) have a central focus on the use of evaluation

for the improvement of management, which is a core principle of his results-oriented evaluation approach. He recognizes and attends to the complexity of managing an organization and views evaluation as an important process to "stimulate effective management" (Wholey, 1983).

In viewing an organization and its constraints, Wholey recognizes that obtaining evaluation information is costly. Thus, the results-oriented evaluation model proposes a four-stage procedure for the "sequential purchase of information." The first of these, *evaluability assessment*, concerns itself with making an initial assessment (or preassessment) of the feasibility of conducting an evaluation with respect to the organization and issues to be examined. This stage of the assessment considers the extent to which evaluation results are likely to be useful for program managers in effecting improved program performance.

The other three stages in the "sequential purchase of information" strategy are *rapid-feedback evaluation*, which focuses primarily on extant and easily collected information; *performance* (or outcome) *monitoring*, which measures program performance, usually in comparison to prior or expected performance; and *intensive evaluation*, which uses comparison or control groups to better estimate the effectiveness of program activities in causing observed results.

Results-oriented evaluation is firmly connected to the potential management benefits (particularly short-term instrumental utilization) of evaluation, placing this model on the use branch of the tree.

Practical Policy Evaluation: Eleanor Chelimsky

"Telling the truth to people who may not want to hear it is, after all, the chief purpose of evaluation" (Chelimsky, 1995). Eleanor Chelimsky is best known for establishing and directing the evaluation unit of the GAO, the largest independent internal evaluation unit in existence. In this role, she developed a unique perspective on evalua-

tion utilization, which we have called *practical policy evaluation*, motivated by the need to inform each of the three branches of the U.S. government and the need for public accountability. Chelimsky has maintained that the evaluation of public policies, programs, and practices is fundamental to a democratic government for four reasons:

(1) to support congressional oversight; (2) to build a stronger knowledge base for policy making; (3) to help agencies develop improved capabilities for policy and program planning, implementation, and analysis of results, as well as learning-oriented direction in their practice; (4) to strengthen public information about government activities through dissemination of evaluation findings. (2006, p. 33)

Thus, practical policy evaluation should generate information for conceptual and enlightenment use, for organizational change and development, and for formative program improvements (i.e., actionable, instrumental use).

This approach does not emphasize individual primary users but rather larger user groups, such as congressional subcommittees, members of particular government agencies, or the general public. Thus, less attention is given to the involvement of particular stakeholders in the evaluation process. Instead, there is a focus on strategies for presenting information so that it is likely to be consumed by members of the target audience(s). Included in the final report are written responses from the organization head (e.g., the head of the federal Title V program) as a way to include transparency and use.

Because of the political nature of government policy evaluation work, practical policy evaluation argues for the most rigorous evaluation study possible given the restraints placed on it. Excellence in procedure and practice can never be compromised in an evaluation, and accuracy and credibility are necessary conditions for evaluation use to occur in political environments. Under Chelimsky's (1997) direction,

to attain excellence, the GAO has devised a wide variety of methods to suit different question types. This flexibility of methodology is a major reason for our placing practical policy evaluation on the use branch.

Context-Sensitive Evaluation: Marvin Alkin

In early years, Marvin Alkin was identified as being focused on evaluation and decision-making concerns. Alkin's early model (1972) had many similarities to the CIPP model, though the primary distinction was Alkin's recognition that *process* and *product* have both summative and formative dimensions. Thus, one could look at process summatively (through program documentation) or at product formatively (through examination of outcomes). The Center for the Study of Evaluation (CSE) model was used as the basis for the CSE Evaluation Kit.

Context-sensitive evaluation, Alkin's most recent view of evaluation theory, employs research in findings on factors associated with attaining evaluation use as a basis for designing essential theory elements. Research factors are embedded within four contexts: the evaluator context, the user context, the organizational/cultural context, and the evaluation context.

The focus of CSE is producing high-quality evaluation procedures that would enhance the use of evaluation to a broad spectrum of influential stakeholders. To attain this goal, it is the evaluator's obligation and personal commitment to help ensure that utilization takes place. Alkin has called this the "evaluator personal factor." *Use* is defined as either instrumental or conceptual in character, but it is specifically focused on the program being evaluated. Alkin has contended that seeking generalizability or potential evaluation "influence" distracts from best serving "this program at this time."

Another distinctive element of Alkin's context-sensitive evaluation approach is the strong emphasis on a thorough examination of program context as a guide for conducting the evaluation. Furthermore, Alkin

has a distinct view on the role of the evaluator in valuing. He strongly rejects the dominant role of evaluators as valuing agents. Instead, he prefers to work with primary users at the outset of the evaluation process to establish value systems for judging potential outcome data. In interactive sessions, he presents a variety of simulated potential outcomes and seeks judgments (values) on the implications of each. This predetermined system forms the basis for judging evaluation findings. Alkin has acknowledged that there are conditions under which it is not tenable to engage intended primary users in this prejudgment process; under such circumstances, he prefers to present evaluation data as factually as possible, without imposing value judgment, unless there are extreme cases that demand that valuing take place.

We have placed context-sensitive evaluation on the use branch in close relationship to Patton's utilization-focused evaluation.

Utilization-Focused Evaluation: Michael Patton

The most prominent theoretical explication of the utilization (or use) extension was developed by Michael Patton (1978, 1986, 1997, 2008). At variance with earlier evaluation and decision-making theorists, Patton has maintained that the evaluator must seek out individuals who are likely to be real users of the evaluation. Patton refers to them as "primary intended users." In Patton's utilization-focused evaluation (UFE), the first step is the identification of intended users. The other four major phases of UFE are the development of users' commitment to the intended focus of the evaluation and to evaluation utilization; involvement in methods, design, and measurement; user engagement in actively and directly interpreting findings and making judgments; and making decisions about further dissemination.

The essence of UFE derives from evaluation utilization research that has demonstrated that a primary element in obtain-

ing use is the "personal factor" (Patton et al., 1977). In other words, the likelihood of an evaluation being utilized is greatly enhanced by the identification of people who have a stake in the evaluation and care personally about the findings it generates—thus, the strong focus on identifying primary intended users. Likewise, the evaluator should be actively involved in developing intended users' commitment to potential utilization by actively and directly engaging them in all stages of the evaluation, thereby fostering buy-in.

To enhance relevance of the evaluation, and thus use, Patton (2008) has urged the evaluator to be "active-reactive-interactive-adaptive." Evaluators need to be *active* in identifying intended users and focusing questions, *reactive* in continuing to learn about the evaluative situation, and *adaptive* "in altering the evaluation questions and designs in light of their increased understanding of the situation and changing conditions" (Patton, 2002, p. 432). *Interaction* is an essential part of each of these three activities. This utilization focus differs from the evaluation and decision-making thrust in not only its concern for identifying intended users but also its willingness to be flexible in modifying prespecified evaluation questions and issues to better serve user needs—and thus, to be adaptive.

The theoretical notions taken up by Patton in UFE have had significant influence on those models further up the branch. Thus, we have placed UFE on the middle portion of the use branch.

Empowerment Evaluation: David Fetterman

David Fetterman is the author of books on empowerment evaluation (1996, 2001), which he has described as a process that encourages self-determination among recipients of the program evaluation. He argues that this process includes training, facilitation, and advocacy, often intending to lead to illumination and liberation (Fetterman, 2001). The goal of empowerment evaluation

is to foster self-determination rather than dependency, to the point where program participants—including clients—essentially conduct their own evaluations. The outside evaluator often serves as a coach or additional facilitator, providing clients with the knowledge and tools for continuous self-assessment and accountability.

Fetterman (1994) has described two general forms of empowerment evaluation that are only subtly different. The main distinction between the two is the extent to which the evaluator participates in the evaluation process. In the first case, evaluators teach program participants to conduct their own program evaluations, training program staff to conduct evaluation studies, thus making them more self-sufficient. In this case, the primary work of the evaluator is to build evaluation capacity. In the second case, the evaluator serves as a coach to facilitate evaluations conducted by others. Using this approach, evaluators enable participants to shape the direction of their evaluations, suggest ideal solutions to their problems, and then take an active role in making social change. Fetterman described all empowerment evaluators as having the potential to serve as “illuminating and liberating facilitators, assisting program participants, freeing themselves from traditional roles and expectations” (p. 306).

In empowerment evaluation, the endpoint is not the assessment of the program’s worth (Fetterman, 1998). Value and worth are not seen as static, and evaluation is viewed as an ongoing process: “Through the internalization and institutionalization of self-evaluation processes and practices, a dynamic and responsive approach to evaluation can be developed to accommodate shifts in populations, goals, value from assessments and external forces” (Fetterman, 1998, p. 382).

Empowerment evaluation seeks to attain utilization as a by-product of enhanced stakeholder empowerment. One form of utilization is conceptual use, and this would certainly be an aspect of empowerment.

Thus, empowerment evaluation is placed on the use branch of the evaluation theory tree.

Practical Participatory

Collaborative Evaluation: J. Bradley Cousins

As the evaluation use branch has grown, collaborative evaluation approaches have been developed from earlier utilization ideas. J. Bradley Cousins’s practical participatory evaluation (Cousins & Earl, 1992; Cousins & Whitmore, 1998) is a further extension of this scholarship. Cousins presumes that user buy-in is essential for utilization. An effective way to achieve buy-in is to have program personnel collaborate in the evaluation process.

Cousins’s ideas flowed from the concern for evaluation utilization expressed in the work of Alkin, Patton, and others. As such, his work continues to reflect the importance of the user personal factor in evaluation and the need for collaboration between the evaluator and program stakeholders to increase the likelihood of utilization. He has extended the notion of “intended primary users” (Patton, 1997), maintaining that organizing groups of intended users as collaborative partners enhances the overall quality, relevance, and subsequent impact of the evaluation. Collaborative evaluations are designed for structured, continued, and active participation of users as opposed to participation in utilization-focused evaluations, which could take on a variety of different forms. Implicit in a collaborative approach is the understanding that utilization takes place within the context of an organization and is best accomplished as a part of organizational development. Cousins and his colleagues have advanced a specific type of collaborative evaluation—practical participatory evaluation (e.g., Cousins & Earl, 1995).

Cousins, together with Earl, defined *practical participatory evaluation* as “applied social research that involves trained evaluation personnel *and* practice-based decision mak-

ers working in partnership” (1995, p. 8). In other words, primary users and evaluators are explicitly recognized as collaborators in the evaluation process. This approach is grounded in a framework that adopts strategies intended to enhance the learning capacity of organizations. Consequently, it is best suited for evaluation projects that “seek to understand programs with the expressed intention of informing and improving their implementation” (Cousins & Earl, 1995, p. 8).

Collaborative evaluation presumes that a commitment to program improvement attracts primary users to the evaluation process, that the partnership within the organization engages them, and that this in turn increases utilization. Responsibilities are shared equally between organization staff and the evaluator. The evaluator trains and supervises agency personnel in technical evaluation skills so that eventually they can coordinate new evaluation efforts. Thus, practitioners learn on the job, and, in time, the evaluator becomes a consultant, responsible only for technical activities, such as instrument design, data analysis, and technical reporting.

Although participatory and empowerment evaluation employ similar practices, their goals are different; to consider them synonymous would be a fundamental misunderstanding of the theories. In practice, the conception of stakeholder involvement that underlies both participatory and empowerment evaluation looks very similar; that is, stakeholders decide on and assist in conducting all aspects of the evaluation, including design, implementation, analysis, and interpretation. But in empowerment evaluation, self-evaluation is promoted as a means for empowering those in a political or emancipatory fashion. This differs from practical participatory evaluation, where the goal of the participatory process is increased utilization through these activities.

Because of these emphases related to utilization, we have placed collaborative evaluation on the use branch of the tree. How-

ever, as we have also noted in other research (Christie, 2003), Cousins has emphasized the use of rigorous methods during his participatory process, with the ultimate goal of increased utilization. To reflect the receptiveness to the ideas reflected in methods theoretical approaches, we have placed collaborative evaluation extending toward the methods branch.

Learning-Oriented Evaluation: Hallie Preskill

The use branch evolves from those theories that focus on use to those that focus on organizational learning and development. These approaches contend that substantial evaluation utilization can occur during the evaluation process and that this can be a valuable tool for transformative learning. The work of Hallie Preskill—with Rosalie Torres and other occasional coauthors—is concerned with providing ongoing learning from evaluation processes and findings leading to evaluation use.

The goal of learning-oriented evaluation is to create transformational learning within an organization through the evaluation process. *Transformational learning*, according to Preskill and Torres (2001), refers to a process where individuals, teams, and even organizations identify, examine, and understand the information needed to meet their goals. They maintain that to aid transformational learning, when an evaluator approaches an organization they should (1) use a clinical approach, (2) span traditional boundaries between evaluator and program staff, and (3) diagnose the organizational capacity for learning.

With this approach, the evaluator provides organizations with more than the technical expertise necessary to conduct an evaluation, such as methods design and data analysis. First, taking a *clinical approach* to evaluation allows for greater focus on the current needs, contextual settings, and historical surroundings of an organization. The approach “is inherently responsive to the needs of an organization and its mem-

bers” (Preskill & Torres, 2001, p. 31) and allows for reflection and the creation of dialogue, which in turn facilitates transformational learning. Furthermore, the evaluator should *span the traditional boundaries* between evaluator and program staff, becoming a facilitator to help guide the learning process that occurs during an evaluation. This process can be used to increase understanding of how the evaluation findings are arrived at, how they can be implemented, and what still needs to be done. The ultimate goal of this process is to create a seamless blend of program work, research, evaluation, and organizational development. Finally, an evaluator needs the ability to *diagnose an organization’s capacity for learning* (Preskill & Torres, 1998); doing so offers a contextual setting to frame evaluation findings and to set realistic goals and expectations for evaluation use. Preskill (2004) has contended that organizational learning can occur when results from the diagnosis are shared because this provides an opportunity for reflection, transformational learning, and, ultimately, utilization.

Learning-oriented evaluation incorporates principles and practices of appreciative inquiry (AI) that build on past successes (and peak experiences) in an effort to design and implement future actions. The philosophy underlying AI is that deficit-based approaches to organizational change are not necessarily the most effective or efficient; that is, when evaluators look for problems, more problems are found; when deficit-based language is used, stakeholders often feel hopeless, powerless, and generally more exhausted. Proponents of AI have found that by reflecting on what has worked well, by remembering topics of study that created excitement and energy, and by using affirmative and strengths-based language, participants’ creativity, passion, and excitement about the future are increased.

The philosophy and principles put forth in learning-oriented evaluation are intended to increase the use of the evaluation processes and findings (Preskill, 2004).

Given the goal of organizational change through evaluation and the role that use has in facilitating the learning process associated with organizational change, we have placed learning-oriented evaluation on the use branch.

Interactive Evaluation: Jean King

Jean King’s interactive evaluation model further extends the use branch’s participatory evaluation models. This approach integrates principles from various participatory, user-focused, and capacity-building approaches. Interactive evaluation presumes long-term engagement with organizations to develop shared understandings and, over time, create structures that will continue to build evaluation capacity (Volkov & King, 2007).

King recognizes that many evaluations are more short term in nature; thus, she has also advanced a particular kind of process—interactive evaluation practice (IEP)—for fostering participation and obtaining use. She defined IEP as “the intentional act of engaging people in making decisions, taking action, and reflecting while conducting an evaluation study” (King & Stevahn, 2012, p. 14). In this procedure, evaluation is viewed as “a process of systematic inquiry designed to provide sound information about the characteristics, activities, or outcomes of a program or policy for a valued purpose” (King & Stevahn, 2012, p. 13). To enhance the possibility of use, it is critical that the evaluator be concerned with creating a participatory environment throughout the evaluation process. This environment must include, among other things, shared meaning of experiences among participants and a great degree of interpersonal and organizational trust.

In IEP, the evaluator’s efforts are focused on engagement in communication and discussion to create shared meaning among participants. It is also important to identify and foster leaders during the evaluation process, as leaders are needed to “attract or recruit people to the evaluation process,

who are eager to learn and facilitate the process . . . and who are willing to stay the course when things go wrong” (King, 1998, p. 64). Trust building is one of the fundamental requirements for a successful participatory evaluation; thus, evaluators should pay close attention to the interpersonal dynamics that transpire during the evaluation process (King & Stevahn, 2012). Without effective interpersonal interaction, the success of the evaluation enterprise is at risk.

The IEP evaluator is part leader, part manager, and part wise counselor. The evaluator is aided in these roles by acting as a reflective practitioner, which “is especially important at key transition points in the study” (King & Stevahn, 2012, p. 19). Thus, IEP focuses on the importance of the interpersonal factor in evaluation leading to evaluation use; hence, its placement on the use branch of the evaluation tree.

Developmental Evaluation: Michael Patton

In recent years, Michael Patton has broadened the scope of the theoretical views put forward in his utilization-focused evaluation to incorporate other theoretical notions and practical applications. His developmental evaluation approach evolved from his work identifying a type of use that takes place as a consequence of “the evaluation process—process use—and not just from evaluation findings” (Patton, 1997, p. 103). Patton (2010) described in *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* how the evaluator becomes part of a program’s design team or management team, fully participating in decisions and in facilitating discussions about how to evaluate. In essence, the evaluator helps to develop the intervention; as a result, we believe that they take on a management-consulting role.

The goal of developmental evaluation is to respond and adapt to social innovators’ information needs in order to meaningfully contribute to program development (Patton, 2016). The focus is on assisting intended

users to “clarify, focus, and articulate what they are trying to do, as they do it” (Patton, 2010, p. 39). In practice, developmental evaluation is context-dependent and does not rely on a prescribed methodology. Rather, key to a developmental approach is frequent engagement between the developmental evaluator and intended users. The intended users should play an active role in determining program goals and making judgments about how the program is progressing. In complement, the evaluator should provide real-time, ongoing insights to guide stakeholder decision making about the direction of social innovation (Patton, 2010). Therefore, the evaluation design should be flexible and incorporate new methods and measures as the program adapts and changes.

Unique to a developmental approach is the positioning of the evaluator as an internal team member. Ideally, the evaluator should be brought into the program development process early to build long-term relationships with the stakeholders. Patton has referred to the developmental evaluator as a “critical friend” who challenges intended users’ assumptions about their program and facilitates evaluative thinking and ways to use data to inform decision making (Patton, 2010). Likewise, the key characteristic of developmental evaluation is continuous learning based on program adaptations through the evaluation process. Developmental evaluation provides not only an opportunity to assist in program development but also strengthens individual and organizational evaluation capacity. Thus, we placed it on the use branch, above the capacity-building approaches.

A FINAL NOTE

We faced two main challenges in writing this chapter. First, the challenge of shifting from discussing theories rather than theorists, as we did in previous versions of the tree. As is the case in other disciplines and fields, it is often impossible to separate the theorist

from the theory—for example, Maslow's hierarchy of needs, and Bronfenbrenner's ecological model. We found it difficult to discuss a theory without also discussing a theorist. This was further complicated by the structure of this book, which connects the theories described in this chapter to the authors of chapters included in this book.

Second, we needed to determine which theories to include on the tree, which also offered particular challenges. These challenges were addressed to a certain extent earlier, in the second chapter, in a description of distinctions between methodologists, evaluation analysts, other evaluation writers, and those who have developed evaluation theories. A further limitation on theory selection was our focus on more general evaluation approaches as opposed to models for evaluation specific to particular fields. Consequently, evaluation models developed by scholars such as Steckler and Linnan (2002) in public health or Astin (1991) in education are not included. It is also important to acknowledge that this book represents the theoretical models advanced by those working mostly in North America, with the notable exception of Cram, a coauthor of Chapter 13 on Indigenous evaluation. Certainly, there are broad theoretical views advanced by authors working in other parts of the world, written in English or in other languages. While not included in this book, theories advanced by those working in Europe, Australasia, Africa, and elsewhere have been discussed to a certain extent in previous editions of the book, in other published articles (e.g., Carden & Alkin, 2012), and by other authors (e.g., Mertens & Wilson, 2018). The theory tree is not intended to provide an exhaustive view of the theory field; thus, we recognize that it does not include theories of relevance to some of today's practitioners and scholars.

Finally, several evaluation theories are represented on the theory tree, but related chapters are not included in the book. These approaches are briefly described in Chapter 22. There are various reasons as to why

chapters were not included, including an invited author declining to write or because the most appropriate chapter author is now deceased.

REFERENCES

- Abma, T. A., & Stake, R. E. (2001). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation*, 92, 7–20.
- Alkin, M. (1972). Evaluation theory development. In C. Weiss (Ed.), *Evaluation action programs* (pp. 105–117). Allyn & Bacon.
- Alkin, M. (Ed.). (1990). *Debates on evaluation*. SAGE.
- Alkin, M. C. (2004). *Evaluation roots: Tracing theorists' views and influences*. SAGE.
- Alkin, M. C. (2012). *Evaluation roots: Tracing theorists' views and influences* (2nd ed.). SAGE.
- Alkin, M., Daillak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* (SAGE Library of Social Research, Vol. 76). SAGE.
- Alkin, M., Hofstetter, C., & Ai, X. (1998). Stakeholder involvement in evaluation. In A. Reynolds & H. Walberg (Eds.), *Evaluation for educational productivity* (pp. 87–113). JAI Press.
- Alkin, M., Koscoff, J., Fitz-Gibbon, C., & Seligman, R. (1974). *Evaluation and decision making: The Title VII experience* (CSE Monograph Series in Evaluation No. 4). Center for the Study of Evaluation.
- Alkin, M., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1–12.
- American Evaluation Association. (2011). *American Evaluation Association statement on cultural competence in evaluation*. www.eval.org/about/competencies-standards/cutural-competence-statement
- Astin, A. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Macmillan.
- Berk, R. A., Boruch, R. F., Chambers, D. L., Rossi, P. H., & Witte, A. D. (1985). Social policy experimentation: A position paper. *Evaluation Review*, 9(4), 387–429.
- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals (Handbook I: Cognitive domain)*. David McKay Company.

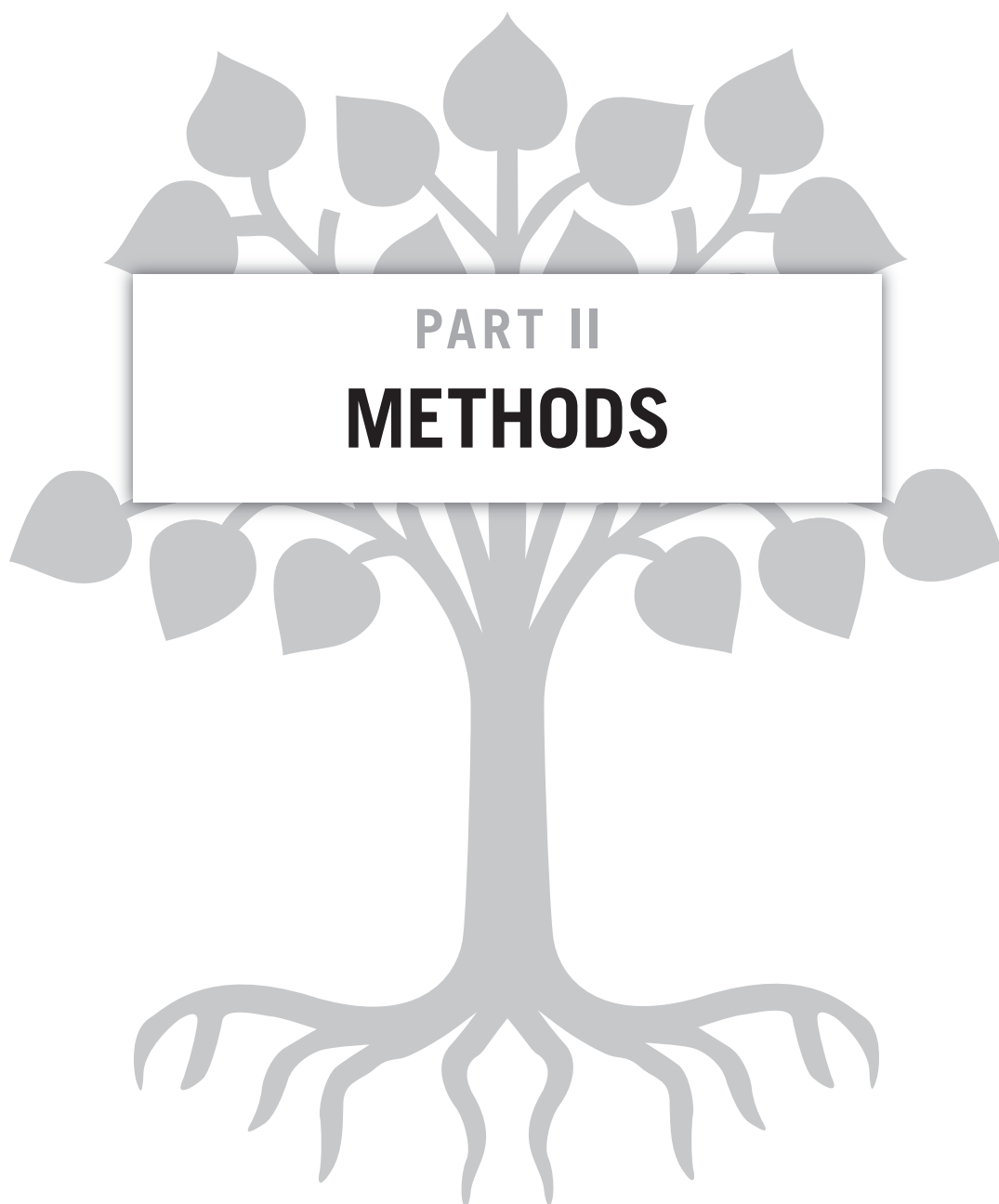
- Boruch, R. F. (1997). *Randomized experiments for planning and evaluation*. SAGE.
- Boruch, R. F., McSweeney, A. J., & Soderstrom, E. J. (1978). Randomized field experiments for program planning, development and evaluation. *Evaluation Quarterly*, 2(4), 655–695.
- Boruch, R., Snyder, B., & DeMoya, D. (2000). The importance of randomized field trials. *Crime and Delinquency*, 46(2), 156–180.
- Bowman-Waapalaneekweew, N. R. (2018). Looking backward but moving forward: Honoring the sacred and asserting the sovereign in Indigenous evaluation. *American Journal of Evaluation*, 39(4), 543–568.
- Braskamp, L. A., Brown, R. D., & Newman, D. L. (1982). Studying evaluation utilization through simulations. *Evaluation Review*, 6(1), 114–126.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312.
- Campbell, D. T. (1975a). Assessing the impact of planned social change. In G. M. Lyons (Ed.), *Social research and public policies: The Dartmouth/OECD conference* (pp. 3–45). Dartmouth College.
- Campbell, D. T. (1975b). “Degrees of freedom” and the case study. *Comparative Political Studies*, 8, 178–193.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Rand McNally.
- Carden, F., & Alkin, M. C. (2012). Evaluation roots: An international perspective. *Journal of Multidisciplinary Evaluation*, 8(17), 102–118.
- Chelimsky, E. (1995). The political environment of evaluation and what it means for the development of the field. *Evaluation Practice*, 16(3), 215–225.
- Chelimsky, E. (1997). The coming transformations in evaluation. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 1–26). SAGE.
- Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation* (pp. 33–55). SAGE.
- Chen, H.-T. (1990). *Theory-driven evaluations: A comprehensive perspective*. SAGE.
- Chen, H.-T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. SAGE.
- Chen, H.-T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7(3), 283–302.
- Chen, H.-T., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, 10(1), 95–103.
- Chen, H. T., & Rossi, P. H. (1992). Using theory to improve program and policy evaluations. Greenwood Press.
- Chouinard, J. A., & Cram, F. (2020). *Culturally responsive approaches to evaluation: Empirical implications for theory and practice*. SAGE.
- Christie, C. (2003). What guides evaluation?: A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–35.
- Christie, C. A. (2007). Reported influence of evaluation data on decision-makers’ actions: An empirical examination. *American Journal of Evaluation*, 28(1), 8–25.
- Christie, C. A., & Fleischer, D. N. (2009). Social inquiry paradigms as a frame for the debate on credible evidence. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 19–30). SAGE.
- Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation*, 40(4), 490–508.
- Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223–326). Rand McNally.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand McNally.
- Cook, T., & Campbell, D. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68(1), 141–180.
- Cook, T. D., & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Quarterly*, 2(1), 5–51.
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4), 397–418.
- Cousins, J. B., & Earl, L. M. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. Falmer Press.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 5–23.

- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. Jossey-Bass.
- Cronbach, L. J., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.
- Diesing, P. (1966). Objectivism vs. subjectivism in the social sciences. *The Philosophy of Science*, 33(1/2), 124–133.
- Durkheim, É. (1966). *The rules of sociological method* (G. F. G. Catlin, Ed.; S. A. Solovay & J. H. Mueller, Trans.; 8th ed.). Free Press. (Original work published 1895)
- Eisner, E. W. (1976). Educational connoisseurship and criticism: Their form and function in educational evaluation. *Journal of Aesthetic Education*, 10(3–4), 135–150.
- Eisner, E. (1985). *The art of educational evaluation: A personal view*. Falmer Press.
- Eisner, E. W. (1991a). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Macmillan.
- Eisner, E. W. (1991b). Taking a second look: Educational connoisseurship revisited. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th Yearbook of the National Society for the Study of Education, Part II, pp. 169–187). University of Chicago Press.
- Eisner, E. W. (1994). *The educational imagination: On the design and evaluation of educational programs* (3rd ed.). Macmillan.
- Eisner, E. W. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Merrill.
- Fetterman, D. M. (1994). Steps of empowerment education: From California to Cape Town. *Evaluation and Program Planning*, 17(3), 305–313.
- Fetterman, D. M. (1996). Empowerment evaluation: An introduction to theory and practice. In D. M. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment and accountability* (pp. 3–48). SAGE.
- Fetterman, D. M. (1998). Empowerment evaluation and accreditation in higher education. In E. Chelmsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 381–395). SAGE.
- Fetterman, D. M. (2001). *Foundations of empowerment evaluation*. Thousand Oaks, CA: SAGE.
- Geertz, C. (1973). *The interpretation of cultures*. Basic Books.
- Greene, J. C. (2000). Challenges in practicing deliberative democratic evaluation. *New Directions for Evaluation*, 85, 13–26.
- Greene, J. C. (2005). Evaluators as the stewards of the public good. In S. Hood, R. Hopson, & H. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 7–20). Information Age.
- Greene, J. C. (2009). Evidence as “proof” and evidence as “inkling.” In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 153–167). SAGE.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. SAGE.
- Henry, G. T. (1995). *Graphing data: Techniques for display analysis* (Applied Social Research Methods Series, Vol. 36). SAGE.
- Henry, G. T. (1996). Does the public have a role in evaluation?: Surveys and democratic discourse. *New Directions for Evaluation*, 70, 3–15.
- Henry, G. T., Julnes, G., & Mark, M. M. (Eds.). (1998). Realist evaluation: An emerging theory in support of practice. *New Directions for Evaluation*, 78, 1–109.
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation’s influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 294–314.
- Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the “Nobody Knows My Name” project, and African educational excellence. *Review of Educational Research*, 78(3), 410–426.
- Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 281–317). Wiley.
- House, E. R. (1991). Evaluation and social justice: Where are we? In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th Yearbook of the National Society for the Study of Education, Part II, pp. 233–247). University of Chicago Press.
- House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. SAGE.
- House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. SAGE.
- House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, 85, 3–12.

- Jamison, D. F. (2019). Asa Hilliard: Conceptualizing and constructing an African-centered pedagogy. *Journal of Black Studies*, 51, 3–15.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). SAGE.
- King, J. A. (1998). Making sense of participatory evaluation on practice. *New Directions in Evaluation*, 80, 57–67.
- King, J. A., & Pechman, E. (1982). *Improving evaluation use in local schools*. National Institute of Education.
- King, J. A., & Stevahn, L. (2012). *Interactive evaluation in practice: Managing the interpersonal dynamics of program evaluation*. SAGE.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5–23.
- Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400–413.
- Kovach, M. (2010). *Indigenous methodologies: Characteristics, conversations, and contexts*. University of Toronto Press.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives, Handbook II: The affective domain*. David McKay Company.
- LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *Canadian Journal of Program Evaluation*, 23(2), 13–31.
- LaFrance, J., Nichols, R., & Kirkhart, K. E. (2012). Culture writes the script: On the centrality of context in Indigenous evaluation. *New Directions for Evaluation*, 135, 59–74.
- Levin, H. M. (2005). Cost-benefit analysis. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 86–90). SAGE.
- Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd ed.). SAGE.
- Madaus, G. F., & Stufflebeam, D. L. (1989). *Educational evaluation: Classic works of Ralph W. Tyler*. Kluwer Academic.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35–57.
- Mark, M. M., Henry, G. T., & Julnes, G. (1998). A realist theory of evaluation practice. *New Directions for Evaluation*, 78, 3–32.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving public and non-profit policies and programs*. Jossey-Bass.
- Marx, K. (1932). *Capital: A critique of political economy. Vol. 1. The process of capitalist production* (F. Engels, Ed.; S. Moore & E. Aveling, Trans.). Charles H. Kerr and Company. (Original work published 1867)
- Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Sage.
- Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20(1), 1–14.
- Mertens, D. M. (2009). *Transformative research and evaluation*. Guilford Press.
- Mertens, D. M., & Hesse-Biber, S. (2013). Mixed methods and credibility of evidence in evaluation. *New Directions for Evaluation*, 2013, 5–13.
- Mertens, D., & Wilson, A. (2018). *Program evaluation theory and practice: A comprehensive guide* (2nd ed.). Guilford Press.
- Mosteller, F. F., & Boruch, R. F. (Eds.). (2002). *Evidence matters: Randomized trials in education research*. Brookings Institution Press.
- National Research Council. (2002). *Scientific research in education*. The National Academies Press.
- No Child Left Behind Act of 2001, Public Law 107–110, 20 U.S.C. § 6319 (2002).
- Patton, M. Q. (1978). *Utilization-focused evaluation* (1st ed.). Sage.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). SAGE.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). SAGE.
- Patton, M. Q. (2002). Utilization-focused evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Vol. 49. Evaluation in education and human services* (pp. 425–428). Kluwer Academic.
- Patton, M. Q. (2008). Advocacy impact evaluation. *Journal of Multidisciplinary Evaluation*, 5(9), 1–10.
- Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.
- Patton, M. Q. (2016). What is essential in developmental evaluation?: On integrity, fidelity, adultery, abstinence, impotence, long-term commitment, integrity, and sensitivity in implementing evaluation models. *American Journal of Evaluation*, 37(2), 250–265.

- Patton, M. Q., Grimes, P., Guthrie, K., Brennan, N., French, B., & Blyth, D. (1977). In search of impact. In C. H. Weiss (Ed.), *Using social research in public policy making* (pp. 141–164). Lexington Books.
- Preskill, H. (2004). The transformational power of evaluation: Passion, purpose, and practice. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences*. SAGE.
- Preskill, H., & Torres, R. T. (1998, November). *Evaluative inquiry as transformative learning* [Conference session]. American Evaluation Association Annual Conference, Chicago.
- Preskill, H., & Torres, R. T. (Eds.). (2001). The learning dimension of evaluation use. *New Directions for Evaluation*, 88, 25–37.
- Rawls, J. (1971). *A theory of justice*. Harvard University Press.
- Robertson, P., Jorgensen, M., & Garrow, C. E. (2004). Indigenizing evaluation research: How Lakota methodologies are helping “raise the tipi” in the Oglala Sioux Nation. *American Indian Quarterly*, 28(3&4), 499–526.
- Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: A systematic approach* (3rd ed.). SAGE.
- Rossi, P. H., Freeman, H. E., & Wright, S. R. (1979). *Evaluation: A systematic approach*. SAGE.
- Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach* (8th ed.). SAGE.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation* (American Educational Research Association Monograph Series on Evaluation No. 1, pp. 39–83). Rand McNally.
- Scriven, M. (1972a). The methodology of evaluation. In R. M. Gagné, R. W. Tyler, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Rand McNally.
- Scriven, M. (1972b). Pros and cons about goal-free evaluation. *Evaluation Comment: The Journal of Educational Evaluation*, 3(4), 1–7.
- Scriven, M. (1974). Evaluation perspectives and procedures. In J. W. Popham (Ed.), *Evaluation in education: Current applications* (pp. 3–93). McCutchan.
- Scriven, M. (1975). Evaluation bias and its control. In G. V. Glass (Ed.), *Evaluation studies review annual: Volume 1* (pp. 101–118). SAGE.
- Scriven, M. (1983). Evaluation ideologies. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 229–260). Kluwer-Nijhoff.
- Scriven, M. (1986). New frontiers of evaluation. *Evaluation Practice*, 7(1), 7–44.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). SAGE.
- Scriven, M. (2001). Evaluation: Future tense. *American Journal of Evaluation*, 22(3), 301–307.
- Scriven, M. (2003). Evaluation theory and metatheory. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation: Part one* (pp. 15–31). Kluwer.
- SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation*, 102, 5–19.
- Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation: Theories of practice*. SAGE.
- Smith, L. T. (2012). *Decolonising methodologies: research and indigenous peoples* (2nd ed.). Zed Books.
- Stake, R. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523–540.
- Stake, R. E. (1974). *Program evaluation, particularly responsive evaluation* (New Trends in Evaluation Report No. 35). Institute of Education, University of Gothenburg.
- Stake, R. E. (1975). Western Michigan University Evaluation Center, Occasional Paper No. 5.
- Stake, R. E. (2000). A modest commitment to the promotion of democracy. *New Directions for Evaluation*, 85, 97–107.
- Stake, R. E. (2001). A problematic heading. *American Journal of Evaluation*, 22(3), 349–354.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 433–466). SAGE.
- Stake, R., Easley, J. A., & Anastasiou, C. J. (Eds.). (1979). *Case studies in science education*. Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign.
- Steckler, A., & Linnan, L. (Eds.). (2002). *Process evaluation for public health interventions and research*. Jossey-Bass.
- Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation*

- models: Viewpoints on educational and human services evaluation* (pp. 117–141). Kluwer-Nijhoff.
- Stufflebeam, D. L. (2000). Lessons in contracting for evaluations. *American Journal of Evaluation*, 21(3), 293–314.
- Stufflebeam, D. (2001). Evaluation models. *New Directions in Evaluation*, 89, 7–98.
- Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. Russell Sage Foundation.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. SAGE.
- Tillman, L. C. (2002). Culturally sensitive research approaches: An African-American perspective. *Educational Researcher*, 31(9), 3–12.
- Trickett, E. J., & Espino, S. L. R. (2004). Collaboration and social inquiry: Multiple meanings of a construct and its role in creating useful and valid knowledge. *American Journal of Community Psychology*, 34(1–2), 1–69.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35(7), 492–501.
- U.S. Government Accountability Office. (n.d.). *About*. Retrieved from www.gao.gov/about.
- Volkov, B. B., & King, J. A. (2007). *A checklist for building organizational evaluation capacity*. <https://wmich.edu/sites/default/files/attachments/u350/2014/organizationevalcapacity.pdf>
- Weiss, C. H. (Ed.). (1972a). *Evaluating action programs: Readings in social action and education*. Allyn & Bacon.
- Weiss, C. H. (1972b). *Evaluation research: Methods for assessing program effectiveness*. Prentice-Hall.
- Weiss, C. H. (1976). Using research in the policy process: Potential and constraints. *Policy Studies Journal*, 4(3), 224–228.
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426–431.
- Weiss, C. H. (1981). Doing research or doing policy [Review of the book *Toward reform of program evaluation: Aims, methods, and institutional arrangements* by L. J. Cronbach]. *Evaluation and Program Planning*, 4(3–4), 397–402.
- Weiss, C. H. (1991). Evaluation research in the political context: Sixteen years and four administrations later. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th Yearbook of the National Society for the Study of Education, Part II, pp. 211–231). University of Chicago Press.
- Weiss, C. H., & Bucuvalas, M. J. (1980). *Social science research and decision-making*. Columbia University Press.
- Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternate route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation*, 26(1), 12–30.
- Weiss, C. H., Murphy-Graham, E., Petrosino, A., & Gandhi, A. G. (2008). The fairy godmother—and her warts: Making the dream of evidence-based policy come true. *American Journal of Evaluation*, 29(1), 29–47.
- Wholey, J. S. (1983). *Evaluation and effective public management*. Little, Brown.
- Wholey, J. S., Hatry, H. P., & Newcomer, K. (Eds.). (2010). *Handbook of practical program evaluation*. Jossey-Bass.



PART II

METHODS

CHAPTER 4

Randomized Field Experiments

Advances in Practice

Robert F. Boruch
Herbert M. Turner, III

INTRODUCTION

In his *Introduction to the Study of Experimental Medicine*, Dr. Claude Bernard (1865/1957) declared, “Man is naturally metaphysical and arrogant and is thus capable of believing that the ideal creations of his mind, which expresses his feelings, also represent reality. From this, it follows that the experimental method is not really natural to him.” Writing in 1865, he would might not have been surprised to encounter sturdy indifference, if not outright hostility, to the idea of randomized controlled trials in the social sector. The perspectives have changed, nonetheless and for the better, in the social and education sciences as they have in medicine.

This chapter focuses on advances in the practice of randomized controlled trials (RCTs) as determined by statistical theory and conjoined with substantive theory about how a program is supposed to work. The benefit of randomized experiments lies in probability theory that assures a fair comparison of the outcomes in a study and

a statistical statement of one’s confidence in the results. In the following, we assume that basic preconditions for mounting a randomized trial are met: The social problem being addressed is serious, the effect of a purported solution is unclear, methods apart from RCTs are inappropriate or undependable, ethical and legal standards can be accommodated, and the evidence from the RCT will be used (Boruch, Allen-Platt, Gerstner, 2019).

The vehicle for characterizing practice here is an assembly of examples from education mainly, and from criminology, health, and other social science sectors. Figure 4.1 illustrates the increase in frequency of trials sponsored by the National Science Foundation and the Institute of Education Sciences since 2002. Charts from different research areas show similar trajectories (e.g., Mosteller & Boruch, 2002).

These core ideas about who or what is randomized and how it is randomized is considered first in what follows. Because operationalization engenders tailoring

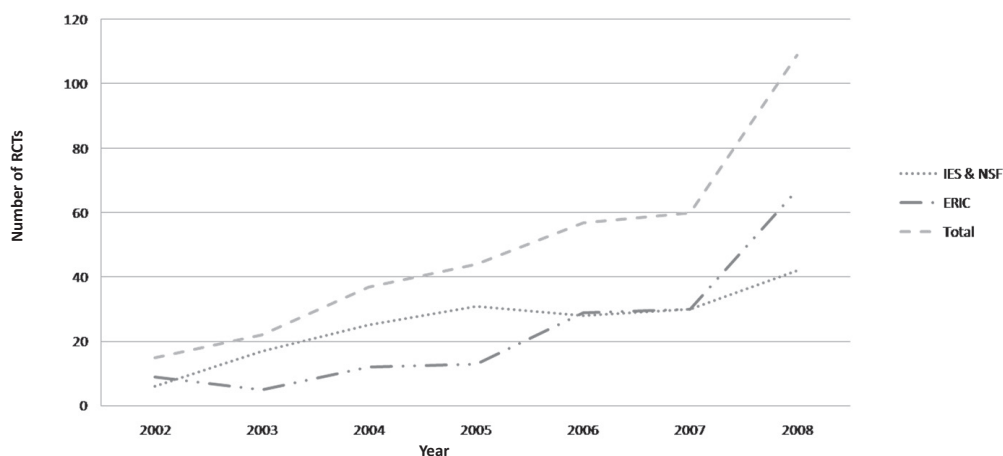


FIGURE 4.1 Trends in the production of randomized trails supported by NSF and IES, based on ERIC. Ad hoc search after 2008 shows similar escalation in frequency.

experimental designs to suit standards of ethics and meeting challenges of political-institutional issues, this too is considered. Replication, reproducibility, and generalization advances are fundamental and summarized briefly as well. Finally, this chapter presents advances in practice that include systematic reviews and meta-analyses of collections of experiments and recent work on failure analysis. This broad framework has its origins in Riecken and Boruch's (1974) *Social Experimentation*.

STATISTICAL THEORY AND ITS OPERATIONALIZATION

We begin this section by considering basic ingredients in a trial such as the units of randomized allocation, and the outcome variables and their frequency of measurement. Moderators and mediators of relationships framed in the experiment are considered next.

Units: Who or What Is Randomized?

A substantive theory may direct attention to individuals and how they behave when subjected to a process or intervention. In the

simplest of randomized trials, the units are assumed to be independent. The theory of random assignment, and indeed much of statistical theory, is based on the independence assumption.

For Sir Ronald Fisher and company, the units of random assignment were independent plants, agricultural plots, beer casks, or cows. The assumption of independence of units was depended on in the early Salk vaccine trials and the streptomycin trials for tuberculosis. This assumption is crucial in contemporary practice such as the international Solidarity Trials for testing vaccines for prevention of COVID-19 by the World Health Organization, the High/Scope preschool program for children, and so on. The assumption is commonplace in laboratory trials in psychology, education, and so forth.

More recently, substantive theories in education, criminology, and other areas have invoked the idea that entire groups, entities, or places ought to be the target of intervention. Within the units of random allocation, the secondary units of observation are not always independent. A case in point is the Tennessee STAR experiments on reducing class sizes so as to gauge the effect of reductions on achievement, something of a

milestone in contemporary history of education trials (see Mosteller, 1995, and references therein). Other illustrations of practice are easy to identify: Middle schools in Uganda have been the targets for educating kids about misinformation in health care (Nsangi et al., 2017), and Zambian schools have been allocated to different regimens to understand the effects of enhanced instructional practice (Seidenfeld & Handa, 2014). Criminologists have deployed sizable “place randomized trials” that focus on innovative approaches to reducing crime through using crime hot spots as randomized units (Braga, Turchan, Papachristos, & Hureau, 2019). The substantive theories of how the interventions are supposed to work are yoked to the statistical assumption of independence of units in practice and theory of randomization.

The statistical theory underlying the randomization of groups or entities has developed apace since the 1970s. For instance, Donner and Klar (2000) produced one of the earliest books on what are nowadays called “cluster randomized trials” in health. Murray’s (1998) “group randomized experiments” focused on entities such as neighborhoods in psychological research. Statistical theory and practice have advanced in the sense that, for cluster trials, intraclass correlation is now well understood, methods for estimation of effect are well developed, and empirical data on magnitudes can be collected (Schochet, 2008, 2020). In education at least, practice has been continuously updated and standardized in that guidance is accessible at the current What Works Clearinghouse (WWC) website¹ and software is accessible at the Society for Research on Educational Effectiveness website.²

Technical advances include easy-to-use software for estimating the statistical power of a trial when substantive theory is sufficient to guess at likely size of an intervention’s effect (Dong & Maynard, 2013). When local conditions determine that the samples must be small in size (say, $n < 10$), Todman and Dugard (2001), among others, provide

software for randomization tests for different experimental designs and illustrations.

Randomization Processes

It’s not always easy to randomize people or entities, and so forth, to different regimens, much less to assure compliance with the regimen. Sir Ronald Fisher himself may have dodged deeper engagement in education experiments following the flaws in conducting his Lanarkshire, Scotland, schools milk trials (Tippett, 1943). The 1980s brought more understanding of the need to tailor processes to suit the setting. A special edition of *New Directions for Program Evaluation*, for instance, included seasoned trialists’ descriptions of various randomization approaches for experiments on welfare and education, treating mental illness, improving tax revenues and judicial processes, and police handling of domestic violence (Boruch & Wothke, 1985). It was flabbergasting to learn that used copies of the issue, initially priced at \$9.95, were being sold on eBay in the late 2010s for about \$400 each.

The approaches to encouraging engagement in trials nowadays include using incentives, both intangible (motivation to improve) and tangible (monetary and otherwise). Prior specification and agreement among stakeholders about eligibility and exclusions are essential for enhancing willingness to participate. The tactics include designing a trial so as to meet ethical or cultural constraints and to stage the implementation of interventions over time, and these topics are considered below.

Longitudinal Study: Are an Intervention’s Effects Sustained Over Time?

Substantive theory of how an intervention is supposed to produce change often focuses on long-term effects rather than a single endpoint. Indeed, noteworthy longitudinal experiments have been done. For instance, the Guatemalan El Caporena trials involved four paired villages randomly assigned to

enriched nutrition versus control regimen, and over 30 years of follow-up with the children of the original participants (Behrman et al., 2009). Seidenfeld and Handa (2014) reported on a 36-month follow-up of a randomized cash transfer program for enhancing children's welfare in Zambia. The High/Scope trial on preschool education has follow-up on its original participants for 50 years (see <https://highscope.org/perry-preschool-project>).

The state of the art in follow-up processes and statistical machinery for analyzing resultant longitudinal data, for characterizing individual or group trajectories, has advanced. See, for instance, Singer and Willet (2003) on theory and its application. The field methods used to engage and sustain contact with individuals over time have themselves been subject to randomized trials (e.g., Lavrakas et al., 2019).

Moderators, Mediators: How Does the Effect Differ Depending on Other Contextual Factors?

The social systems in which many evaluators work are invariably complex. What happens inside the intervention, in the environment, and to subgroups may of course be theoretically important on substantive grounds. Thus, theory of statistical design of a trial and analysis must then take this into account.

As a consequence, “interactions” that characterize how one factor may have a multiplicative effect with another factor to produce an effect that is different from a simply additive effect of either one are important. The statistical model and analysis of variance for interactions have been used for decades. The work is reflected in practice, for example, in reports on trials in education such as Matsumura, Garnier, and Spybrook (2013) on literacy coaching.

The idea of “mediation” suggests that a variable that is intrinsically associated with the randomized intervention may mediate the effect of the intervention on an ultimate outcome. It relies less on randomization to

provide insight than it does on statistical modeling. For instance, see Farwah Fatima (2020) on Ghanaian trial data on teachers and references therein. Advances in this area are summarized along with lots of examples by Hayes's (2018) book on mediators, moderators, and conditional processes.

ETHICS, LAW, AND EXPERIMENTS

Ensuring that human rights are respected in RCTs, and in social research more generally, is imperative. To ensure ethical propriety, substantial progress has been made in developing institutional standards, government policy, law, and the statistical design of trials. Institutional review boards, for instance, have developed since the 1960s, in the United States, the United Kingdom, Canada, and other countries to oversee the ethics of a study, including experiments. Professional societies such as the American Educational Research Association (AERA), the American Sociological Association (ASA), the American Psychological Association (APA), and others have produced codes of ethics to guide practice. These standards usually emphasize individuals' informed consent, although government mandates at times may supersede individual consent. For instance, law and serious cultural considerations can also be construed as embodiments of social ethics that trialists must consider. The legality of RCTs for improving government programs is reflected in at least three decades of debate and practice in welfare, labor, and employment areas (Gueron & Rolston, 2013). More recent justification for the United States hinges on the need for dependable evidence in policy Foundations for Evidence Based Policy Act of 2018 and relevant regulations (U.S. Office of Management and Budget, 2020).

The emergence of cluster randomized trials has engendered ethics issues beyond individual consent. For instance, who gives informed consent for random allocation to different regimens on behalf of an entire

hospital, village, or neighborhood? Weijer and colleagues (2011) have produced guidelines that are being adopted in the health area. It is not clear, however, if the guidelines have become part of normal practice in the social and economic areas, especially in low- and middle-income countries.

Experimentalists have developed ways to tailor randomization so as to conform to ethical or legal requirements. Delaying treatment for control groups, for instance, is sometimes a feasible and not uncommon approach when all eligible individuals must receive the purportedly beneficial program. Variations on the idea include what Glennester and Takavarasha (2013, pp. 170–171) call “rotation designs” in which, for example, schools were randomly assigned such that “in the first year, Group A would receive tutors in grade 3 and Group B would get them in grade 4 . . . in the second year, Group A would receive tutors for grade 4 and Group B receive tutors for grade 3.” Randomization at the cutoff point/margin of a regression discontinuity is used in other designs. As well, “trap door” provisions and criteria that identify exclusions *prior* to randomization can resolve some issues (e.g., the boss’s relatives are separated from the group randomized ahead of time and are guaranteed service, regardless of the fact that its benefits are unknown).

Step wedge designs that ensure that all targets get treated over time are newer but not uncommon when ethics and political-institutional factors demand complete population coverage of a program (see the next section). The program is deployed in stages to a population such that, at the first stage, a random subset of the target areas, entities, or people is randomly selected to receive the program and the remainder serves as a control condition. At the second stage, another random subset is selected to receive the program and, again, the remaining nonrecipients serve as the controls. The sequence is repeated until all subsets receive the program. Recent illustrations include training of police department members in

deescalation of volatile encounters by Engel, Corsaro, Isaza, and McManus (2020).

ARE EXPERIMENTS REPLICATED AND RESULTS GENERALIZED?

If experimental procedures and the intervention being studied cannot be replicated, the product is more art than science. If analysis of the data resulting from a trial cannot be reproduced, the original results or reporting on them are suspect. The dispassionate observer values results of a trial that can be generalized over time or institutional space, but this is a hard target.

To clarify matters, a task force of the Society for Prevention Research has laid out definitions, provided examples, and discussed persistent issues in this area (Valentine et al., 2011). The notion of conceptual replication, involving a second testing of the major ingredients of an earlier trial, is among the explained ideas. This guidance has been recognized as useful in the context of criminological research (Farrington et al., 2019). In education, we now have similar guidance from government funders such as the National Science Foundation (NSF) and Institute of Education Sciences (IES) on replication and reproducibility. The publications provide nice illustrations of conceptual replications in early education childhood trials that have been sponsored by the agencies (NSF & IES, 2018).

Replications are common in the health sector. Indeed, they are required by law in approval of drugs by the U.S. Food and Drug Administration, for instance. However, they are not nearly as common in the education and other social sciences. The reasons for small incremental change in this sector lie with limitations on resources, among other things. Operationalizing the idea is gaining ground slowly and incrementally.

Absent an empirical replication of a trial, one may nonetheless wish to predict whether results of a trial in one setting, or an assembly of settings, can be generalized to other

settings. Consequently, statistical methods and software have been developed to help one understand whether the features of one experimental setting (e.g., neighborhood demography) can be matched against the features of other schools' settings (Tipton & Miller, 2016). The software is available on the Society for Research on Educational Effectiveness (SREE) website. Furthermore, statistical models have been built to facilitate limited generalizations. Chan (2018), for instance, employed statistical theory and archived data to learn how the program impact estimates from a large cluster randomized trial in Indiana could be expected to generalize to schools in the entire state or to classes of other schools within the state.

MANAGEMENT

Time frames, funding sources, and contributors are usually identified in reports on experiments. Details of resolving managerial and political-institutional issues, however, are rarely published. This is partly because the requisite knowledge and approaches to enhanced practice have a proprietary character. Scientists' competition for funding, at times, may trump knowledge sharing of this kind.

Some contemporary trialists nonetheless have advanced our understanding of how to deploy experiments in the field. At the systems level, for instance, Gueron and Rolston's (2013) *Fighting for Reliable Evidence* covers experience in mounting welfare experiments in many state and local jurisdictions. It is a well-informed and detailed handling of real-world challenges to experimentation in politically complex and changing environments.

For project-level studies in low- and middle-income countries, Glennerster and Taka-varasha (2013) demonstrated how temporal, logistical, and resource challenges have been resolved in Kenya deworming experiments and others. Bloom (2005) tells us how cluster randomized trials have been successful. Peck's (2020) volume is brief, but it is instruc-

tive about operationalizing randomization plans and practical considerations.

More recently, experimental designs have been operationalized to learn about effects of different ways to program and disseminate results. Brown and nearly 20 colleagues formed a working group under the auspices of the National Institutes of Health to produce an instructive report on the topic (Brown et al., 2017). The latter is studded with examples of emerging practice to test management strategies. For instance, cluster randomized trials on new ways to implement health care standards, comparisons of effects of technology-based patient-oriented versus physician-oriented approaches to care management, alternative strategies for foster care, and many others are explained at least briefly. The emphasis is on collaboration among evaluators and the organizations or communities that must change in translating substantive research to practice.

SYSTEMATIC REVIEWS, META-ANALYSIS, AND QUALITY CONTROL

The idea of systematically reviewing a collection of randomized trials and summarizing the effect sizes in the social sciences was advanced early by Glass (1976), planted firmly in the professional's imagination through compelling stories by Hunt (1997), and extended methodologically by specialists. The advances in this sector are given in fine detail in the third edition of a handbook on the topic edited by Cooper, Hedges, and Valentine (2019).

Operationally defined, a systematic review is the application of procedures that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a particular topic. The steps of the review process involve formulating review questions, defining the inclusion criteria, locating and selecting studies, assessing study quality, and coding the data. Meta-analysis may also be part of the process (Chalmers, Hedges, & Cooper, 2002). A meta-analysis is a statistical process attaching greater weight to study

reported effects with large samples and lesser weight to study reported effects with smaller samples to maximize the statistical confidence in the pooled effect (Borenstein, Hedges, Higgins, & Rothstein, 2009).

The growth in RCTs has resulted in systematic efforts to build databases on RCTs and to construct registers that facilitate tracking their progress. See, for instance, the Manpower Demonstration Research Corporation (MDRC; 2021) database on RCTs in higher education, the website for *clinicaltrials.gov*, and Spybrook, Anderson, and Maynard (2019) on education stud-

ies. These parallel the growth in systematic review organizations that try to identify what works, primarily by assessing the internal validity of experiments.

Before 2002, only three such organizations routinely operated in the United States: (1) the General Accounting Office (GAO), (2) Blueprints for Violence Prevention (BVP), and (3) the Center for Evaluation and Research Methods (CERM). As of 2020, at least 14 systematic review clearinghouses are operating in the United States (Table 4.1). These clearinghouses are sponsored by federal and state governments, foundations,

TABLE 4.1. The Most Visible and Active Systematic Review Clearinghouses in the United States

Clearinghouse	Sponsor	Website
Pathways to Work	U.S. Department of Health and Human Services	https://pathwaystowork.acf.hhs.gov
Best Evidence Encyclopedia	Johns Hopkins University School of Education	www.bestevidence.org
Blueprints for Healthy Youth Development	University of Colorado Boulder: Institute of Behavior Science	www.blueprintsprograms.org
California Evidence-Based Clearinghouse for Child Welfare	California Department of Social Services	www.cebc4cw.org
Clearinghouse for Labor Evaluation and Research	U.S. Department of Labor	https://clear.dol.gov
Crime Solutions	U.S. Department of Justice	https://crimesolutions.ojp.gov
Cross State High School Collaborative	Seven State Joint Effort	www.hsredesign.org
Evidence for ESSA	Johns Hopkins University School of Education	https://education.jhu.edu/2020/02/evidence-for-essa
Ohio's Evidence-Based Clearinghouse*	Ohio Department of Education	https://essa.chrr.ohio-state.edu/getting-started
Results First Clearinghouse*	PEW Charitable Trust	www.pewtrusts.org/en/research-and-analysis/issue-briefs/2014/09/results-first-clearinghouse-database
Self-Sufficiency Research Clearinghouse	U.S. Department of Health and Human Services	https://selfsufficiencyresearch.org
Social Programs That Work	Arnold Ventures	https://evidencebasedprograms.org/about-us
Teen Pregnancy Prevention Evidence Review	U.S. Department of Health and Human Services	https://tpevidencereview.youth.gov
What Works Clearinghouse	U.S. Department of Education's Institute of Education Sciences	https://ies.ed.gov/ncee/wwc
What Works in Reentry Clearinghouse	U.S. Department of Justice	https://whatworks.csgjusticecenter.org

Note. An asterisk (*) denotes a clearinghouse of clearinghouses. The list is suggestive and not exhaustive.

universities, and philanthropies. One common denominator is the systematic review and meta-analysis of both RCTs and quasi-experimental designs (QEDs), and a consistent set of standards applied during the review process. As well, many include the application of Rubin's theory of causal inference (Rubin, 1974) and/or Campbell and Stanley's classic threats to internal validity (Campbell & Stanley, 1963).

The WWC is among the most visible and active clearinghouses in the United States. Founded in 2002 by the IES, and based in part on the International Campbell Collaboration's early work, the WWC is staffed by over 300 individuals collaborating among six large and small policy research firms operating in for-profit and nonprofit sectors in the United States. Since its creation, WWC reviewers have applied the WWC procedures and standards to review over 15,000 U.S. studies as reported in the education and social science literature.³ These reviews have important methodological lessons that intervention researchers (and developers) can use to improve the design, conduct, and reporting of RCTs and QEDs.

For example, more than half of all studies (58%) identified for a WWC review were ruled ineligible. The primary reason was the study abstract or summary indicated that the study aimed to estimate an intervention's effectiveness, but lacked a comparison group. This lack of a counterfactual undermined the validity of any assertion about the intervention's effect.⁴ This "counterfactual" assessment is fundamental to applying Rubin's theory of causal inference in practice (Shadish, Cook, & Campbell, 2001). The valid empirical identification of whether an intervention works requires a control group formed by random (RCT) methods or a comparison group formed by nonrandom (QED) methods. The former is a stronger counterfactual than the latter, from a Clearinghouse perspective, as expressed in the *WWC Standards Handbook 4.1*.

Fundamental ingredients of WWC reviews include explicit standards for a "rigorous"

study to identify what works, standards that are articulated in detail, summarized, and applied. These WWC Evidence Standards outlined below can be summarized using Campbell and Stanley's (1963) classic design diagrams for RCTs and QEDs, which themselves are extended neatly in Reichardt's (2019) book.

Group Formation

How groups are formed is at the heart of experimental design. Making valid causal inferences about an intervention's effects requires either a (1) control group formed by random assignment or a (2) comparison group formed by using nonrandom methods. A "rigorous" study must have this counterfactual group to validly identify and quantify an intervention's effect (*WWC Standards Handbook 4.1*, pp. 4–9)

Outcome Variable Requirements

How outcomes are measured and operationalized through data collection can determine whether a study meets the evidence standards. A "rigorous" study designed to make valid causal inferences about an intervention's effect uses reliable and valid outcomes. These outcomes are collected in the same way for study groups and are not overaligned with the intervention. For example, an outcome can be overaligned with an intervention by using a subset of the items on the measure as part of the intervention, making it invalid in the sense of "teaching to the test" (*WWC Standards Handbook 4.1*, pp. 83–85).

Confounds

How groups are formed should not introduce a *confound*, a study factor or characteristic correlated with an outcome that aligns with one study group but not the other. A study factor correlated with the outcome and aligned this way cannot have its relationship with the outcome disentangled

from the intervention's effect on that outcome. A "rigorous" study designed to make causal valid causal inferences does not introduce confounds through design (*WWC Standards Handbook 4.1*, pp. 85–86).

Attrition

Attrition can be defined as any study participants who are randomized to conditions and missing an outcome. The "who are randomized" aspect of the definition is one reason intervention researchers seek consent of individuals or groups of individuals before randomization; individuals who refuse consent before randomization do not count as attrition; by definition, they have not been randomized. When the combination of overall and differential attrition is high, defined by the WWC's threshold, the attrition disrupts the long-run equating properties of randomization. The high-attrition RCT is methodologically considered a QED and reviewed as such.

Baseline Equivalence

When the study is a high-attrition RCT or QED, the intervention and control (or comparison) groups must be equivalent at baseline on a preintervention measure of an outcome in the analytic sample. The emphasis here is on the analysis sample rather than the baseline sample, so that preexisting group differences due to measured factors (other than group assignment) can be empirically identified and controlled. For instance, this involves using a covariate adjustment in the statistical model or difference-in-difference adjustment if relevant assumptions are met.

Missing Data

When baseline (pretest) or outcome (posttest) data are imputed, the imputation method should not introduce bias. Imputing missing data at baseline or outcome is easy with the wide range of available software. However, it is also easy to introduce bias

during the imputation process, as it is easy to implement the imputation process incorrectly. Specific guidance on what methods are appropriate, how they should be used, and how studies that use them should be reviewed can be found in the *WWC Standards Handbook 4.1*, pages 34–43, and are summarized in Table II.6, pages 37 and 38.

Like the WWC, most of the systematic review clearinghouses that are listed in Table 4.1 published their review standards on their websites. The six standards presented earlier (1) reflect the fundamental methodological standards common across most of these clearinghouses, (2) summarize the fundamental methodological criteria used to determine whether a study meets standards (with or without reservations), and (3) embody those general characteristics of a "rigorous study" that should be favorably reviewed if identified by a clearinghouse for review. In sum, a "rigorous" study is not necessarily a statistically sophisticated or methodological complex one. Instead, it is a study that controls for factors other than the intervention, preferably by design, that explains group differences on the outcome.

TOWARD A SCIENCE OF FAILURE ANALYSIS

Despite advances in substantive theory and practice, and the brainpower and resources invested in testing a novel intervention, one may discern no differences among regimens being compared. So-called null findings are not unusual, nor should they be unexpected. Work by Boruch and Ruby (2015), for instance, emphasized the need to learn from such studies and put the matter into larger scientific context. A special 2019 issue of *Education Researcher* included excellent papers by industrious trialists on the topic. Allen-Platt, Gerstner, Boruch, and Ruby (2021) summarized reasons given by trialists for failure to detect effects in over 50 published studies. Beyond these, Steiner, Wong, and Anglin (2019) and others have

advanced thinking and empirical research on replication and failures to replicate experiments. These works auger well for better learning from failure and for designing both interventions and trials.

CONCLUDING REMARKS

L. H. C. Tippett, former president of the Royal Statistical Society, ran off his rails a bit in the 1940s, declaring, “In most fields that are subject to statistical inquiry, the opportunities for controlled experiments are very few” (1943, p. 132). Contemporary experience makes it clear that randomized trials are increasing in frequency, diversity of application, and the tailoring of study designs to suit ethical, societal, and policy dictates. It is also clear that the experimental designs must focus on the testable features of the substantive theory (i.e., those features that are embodied or operationalized in the interventions being investigated). Our own forecast is that the dependence on randomized trials and their underlying probability theory will flourish and, moreover, that the products of experiments will be better understanding of what works and what does not and for whom, when, and why.

NOTES

1. Available at <https://ies.ed.gov/ncee/wwc/resources/resourcesforreviewers>.
2. Available at www.sree.org/software.
3. The number of studies reviewed is based on our independent analysis of the records retrieved from the WWC Review of Individual Studies database on October 1, 2020. This database is publicly accessible on the WWC website at <https://ies.ed.gov/ncee/wwc/reviewed-studies#/onlystudieswithpositiveeffects:false,setnum ber:1>.
4. There are other reasons that a study was ineligible for review (e.g., the intervention was not relevant to the review), but the lack of comparison group is the most frequent reason.

REFERENCES

- Allen-Platt, C., Gerstner, C.-C., Boruch, R., & Ruby, A. (2021). Toward a science of failure analysis. *Review of Research in Education*, 45, 223–252.
- Behrman, J. R., Calderon, M. C., Preston, S. H., Hoddinott, J., Martorell, R., & Stein, A. D. (2009). Nutritional supplementation in girls influences the growth of their children: Prospective study in Guatemala. *American Journal of Clinical Nutrition*, 90(5), 1372–1379.
- Bernard, C. (1957). *An introduction to the study of experimental medicine* (H. C. Green, Trans.). Dover. (Original work published 1865)
- Bloom, H. S. (2005). *Learning more from social experiments: Evolving analytic approaches*. Russell Sage Foundation.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. Wiley.
- Boruch, R., & Ruby, A. (2015). To flop is human: Can we invent better scientific approaches to anticipating failure to meet expectations in randomized controlled trials and to learn from failure? In S. Kosslyn, R. Scott, & N. Pinkerton (Eds.), *Emerging trends in the behavioral and social sciences* (pp. 1–16). Wiley.
- Boruch, R. F., Allen-Platt, C., & Gerstner, C.-C. (2019). To randomize or not: That is the question. *New Directions for Evaluation*, 163, 73–82.
- Boruch, R., & Wothke, W. (Eds.). (1985). *Randomization and field experimentation* (New Directions for Program Evaluation, Vol. 28). Jossey-Bass.
- Braga, A. A., Turchan, B. S., Papachristos, A. V., & Hureau, D. M. (2019). Hotspot policing and crime reduction. *Journal of Experimental Criminology*, 15, 289–311.
- Brown, C. H., Curran, C., Palinkas, L. A., Aarons, G. A., Wells, K. B., Jones, L., . . . Cruden, G. (2017). An overview of research and evaluation designs for dissemination and implementation. *Annual Review of Public Health*, 38, 1–22.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.
- Chalmers, I., Hedges, L., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation and the Health Professions*, 25, 12–27.
- Chan, W. (2018). Applications of small area

- estimation to generalization with subclassification by propensity scores. *Journal of Educational and Behavioral Sciences*, 43(2), 182–224.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.
- Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect size and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6(1), 24–67.
- Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomized trials*. Arnold.
- Engel, R. S., Corsaro, N., Isaza, G. T., & McManus, H. D. (2020). *Examining the Integration of Communications, Assessment, and Tactics (ICAT) de-escalation training for the Louisville Metropolitan Police Department: Initial findings*. International Association of Chiefs of Police.
- Farrington, D., Losel, F., Boruch, R., Gottfredson, D., Mazerolle, L., Sherman, L., & Weisburd, D. (2019). Advancing knowledge about replication in criminology. *Journal of Experimental Criminology*, 15, 373–396.
- Fatima, S. (2020). *Explaining teacher performance in Ghana*. Unpublished doctoral dissertation, University of Pennsylvania.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8.
- Glennester, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.
- Gueron, J., & Rolston, H. (2013). *Fighting for reliable evidence*. Russell Sage Foundation.
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis*. Guilford Press.
- Hunt, M. M. (1997). *How science takes stock: The story of meta-analysis*. Russell Sage Foundation.
- Lavrakas, P., Traugott, M. W., Kennedy, C., Holbrook, A. L., de Leeuw, E. D., & West, B. T. (Eds.). (2019). *Experimental methods in survey research*. Wiley.
- Manpower Demonstration Research Corporation. (2021). *THE-RCT database*. www.mdrc.org/project/higher-education-randomized-controlled-trial#overview
- Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2013). Literacy coaching to improve reading achievement: A multi-level mediation model. *Learning and Instruction*, 25, 35–48.
- Mosteller, F. (1995, Summer). The Tennessee study of class size in the early grades. *The Future of Children*, pp. 113–127.
- Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in education research*. Brookings Institute.
- Murray, D. M. (1998). *Design and analysis of group randomized trials*. Oxford University Press.
- National Science Foundation & Institute of Education Sciences. (2018). *Companion guidelines on replication and reproducibility in education research*. Authors.
- Nsangi, A., Semakula, D., Oxman, A. D., Austvoll-Dahlgren, A., Oxman, M., Rosenbaum, S., . . . Sewankambo, N. K. (2017). Effects of the informed health choices school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects, 1-year follow-up: A cluster randomized trial. *Lancet*, 390, 374–388.
- Peck, L. (2020). *Experimental evaluation: Design for program improvement*. SAGE.
- Reichardt, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford Press.
- Riecken, H. W., & Boruch, R. F. (1974). *Social experimentation: A method for planning and evaluating social intervention*. Academic Press.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.
- Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics*, 33(1), 62–87.
- Schochet, P. Z. (2020). Analyzing grouped administrative data for RCTs using design-based methods. *Journal of Educational and Behavioral Statistics*, 45(1), 32–57.
- Seidenfeld, D., & Handa, S. (2014). Zambia's Child Grant Program: 36 Month Impact Report. American Institutes for Research.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis*. Oxford University Press.
- Spybrook, J., Anderson, D., & Maynard, R. (2019). The registry of efficacy and effective-

- ness studies. *Journal of Research on Education Effectiveness*, 12(1), 5–9.
- Steiner, P. M., Wong, V. C., & Anglin, K. (2019). A causal replication framework for designing and assessing replication efforts. *Zeitschrift für Psychologie*, 227(4), 280–292.
- Tippett, L. H. C. (1943). *Statistics*. Oxford University Press.
- Tipton, E., & Miller, K. (2016). *The Generalizer*. www.thegeneralizer.org
- Todman, J. B., & Dugard, P. (2001). *Single case and small -n experimental designs: A practical guide to randomization tests*. Erlbaum.
- U.S. Office of Management and Budget. (2020). *Memorandum for Heads of Federal Department and Agencies: Phase 4 Implementation of the Foundations for Evidence-Based Policy Act of 2018: Program Evaluation Standards and Practices*. Author.
- Valentine, J. C., Biglan, A., Boruch, R. S., González Castro, F., Collins, L. M., Flay, B. R., . . . Schinke, S. P. (2011). Replication in prevention science. *Prevention Science*, 12, 103–117.
- Weijer, C., Grimshaw, J. M., Taljaard, M., Binik, A., Boruch, R., Brehaut, J. C., . . . Zwarenstein, M. (2011). Ethical issues posed by cluster randomized trials in health research. *Trials*, 12, 100.

CHAPTER 5

Multiple Paths to Evaluation Influence and Social Betterment

Melvin M. Mark
Gary T. Henry

INTRODUCTION

In this chapter, we describe the key components of our theory of evaluation practice. Our theory does not specify a single approach to evaluation. Rather, it presents a contingent approach, tailored to the specific context in which an evaluation is planned and implemented, and to the specific pathways available in that context for improving social conditions, a goal we refer to as *social betterment*. Our approach is premised on the idea that evaluation exists to improve the accuracy of information about such things as the description of programs, the people they serve, and the outcomes they produce. This is the reason evaluation rests on the application of systematic research methods.

A second key premise of our approach is that evaluation exists to help move us toward social betterment. Understanding the context of the evaluation in specific ways can guide the planning of an evaluation that is more likely to move society toward improving social conditions. Evaluation influence

links an evaluation to its contribution to social betterment, with evaluation purpose based on context, as we describe below. Our theory or approach to evaluation is admittedly complex, befitting the complex world that evaluators seek to influence. We begin with our rationale for evaluation to address the flaws inherent in the natural sensemaking in which humans are predisposed to engage.

EVALUATION AS ASSISTED SENSEMAKING

Humans by nature strive to make sense of the world around them. Indeed, the sense-making motivation is so strong that people often see patterns when none exist. In making sense of the world, people also make evaluative judgments, seemingly automatically. As practicing evaluators know, but as evaluation theories sometimes seem to ignore, stakeholders do not hold off on making evaluative judgments until the findings from an evaluation are available. However,

the informal and everyday evaluations that stakeholders make are often inaccurate, in part because even those working closely with a program's clients typically cannot account accurately for how well the clients would have done in the absence of the program or for factors such as the effects of clients' self-selection into a program. Moreover, stakeholders' personal investment in the program may stimulate biases, such as greater attention to and recall of successful cases, or self-interest in maintaining the status quo.

In short, although humans naturally make evaluative judgments, these judgments are susceptible to limited information and potential biases. This observation motivates the idea of *evaluation as assisted sensemaking* (Mark, Henry, & Julnes, 2000). The fundamental idea is that evaluation can contribute by extending, enhancing, and checking the natural sensemaking that people engage in about programs, policies, or other entities that are the subject of evaluations. Evaluators have at their disposal a set of methodological tools (e.g., randomized experiments, case studies) from the social sciences that can correct, at least in part, the biases and limitations on information that otherwise can affect everyday sensemaking capabilities. These tools can provide stakeholders with more accurate information about a program and its outcomes, thereby increasing the capacity of stakeholders to understand and undertake well-informed actions with regard to the evaluated program or policy.

The notion of evaluation as assisted sensemaking, while perhaps not a stunning one to thoughtful members of the evaluation community, nevertheless has what we see as nontrivial implications. First, it locates evaluation practice squarely in a broader context of human judgment and action processes. It fits well with the idea of organizational decision making as complex and messy, with information often competing with values and politics (Weiss, 1988). But it can also accommodate narrower user-

focused approaches in which the evaluator serves the information needs of specific, interested stakeholders to facilitate use (e.g., Patton, 2008). Moreover, the organizational contexts in which sensemaking about programs takes place vary, so the demands on and approaches to evaluation should vary accordingly. Second, an evaluation, however technically proficient it may be, is of little value if the findings are not shared in ways that connect with the natural sensemaking processes of relevant parties. This has implications not only for the dissemination and presentation of evaluation findings but also for the potential array of evaluator roles related to evaluation influence—a topic to which we return later. Third, while social science research methods can eliminate or reduce bias and other limitations of natural sensemaking, they are neither a panacea for all limitations on sensemaking nor are they so limited in their benefits that their use should be disregarded by evaluators or evaluation consumers.

The careful application of quantitative social science methods can improve the accuracy of the information that evaluators provide to assist sensemaking, with such benefits often being described in terms of validity and reliability. Attention to rigorous research design, including quasi-experimental and experimental designs, can bring the promise of more accurate estimates of program impacts. Attention to measurement issues, such as the quality and relevance of measures used to describe program outcomes, can also improve the accuracy of the information provided through evaluations. Careful application of probability sampling can produce an accurate statistical picture of a program's target population, including their characteristics and outcomes, and coverage of intended beneficiaries. Moreover, advanced quantitative methods can offer insights that go deeper, such as mediational analyses that can increase understanding about the underlying mechanisms through which programs produce their effects (Henry, Pham,

& Zimmer, 2020) and tests of moderated relationships that assess whether effects are equitably distributed to subgroups of beneficiaries (Zimmer, Henry, & Kho, 2017). However, we view quantitative methods as *an* important set of tools in the toolkit, not the only set. Our view, shared with many others, is that the goal for assisted sensemaking in evaluation is to get the most accurate answer for the most pressing questions in a particular circumstance. But how do we propose to select the most pressing questions from all the interesting or at least possible ones?

SOCIAL BETTERMENT AS THE GUIDING STAR

Why does systematic evaluation exist? There are several ways of answering this question. One kind of response highlights the role of evaluation as assisting the natural but all too frequently biased sensemaking on the part of stakeholders. By itself, this does not adequately motivate evaluation, except perhaps by advocacy of accuracy for the sake of accuracy. Another, more compelling response highlights the ultimate goal of evaluation. For example, should the ultimate goal of evaluation—its guiding star and the key criterion for judging success of evaluations itself (Miller & Campbell, 2006)—be utilization, or empowerment, or inclusion, to take a few examples represented elsewhere in this book? We believe that the ultimate guiding star for evaluation should be *social betterment*. Admittedly, social betterment can be a distal concept, but it focuses evaluators' attention on their role in improving social conditions. We envision evaluation as a stimulus for better deliberative decision-making processes, especially in the public sector of democratic societies.

In other words, evaluation is a social intervention, much like the policies and programs that are the subject of evaluation (Henry & Mark, 2003; Mark & Henry, 2004). As such, we have advocated for the use of systematic values inquiry during the evaluation process, often as a precursor to

an evaluation, to broaden and deepen the discussion of program outcomes. We use the term *values inquiry* to suggest that evaluators may need to use systematic research methods to understand better the values that should be infused into an evaluation. This would include getting a better understanding of which of the intended objectives for a program are most valued by various constituencies, as well as unintended consequences or negative side effects that are feared by important constituencies. For example, Henry (2002) surveyed parents, teachers, program administrators, and the general public in an effort to understand each group's view of the most important outcomes that could result from a state-wide prekindergarten (PreK) program. The highest priority outcomes were different for these groups, which required an array of outcomes to be measured for the evaluation. Among the benefits of systematic values inquiry is that it enhances transparency about stakeholder input, which otherwise is often opaque.

The somewhat stilted term *social betterment* has nonetheless been widely adopted and was chosen intentionally to convey several specifics. First, it is distinguished from social progress, which has been taken to mean a steady, inevitable improvement of social conditions. Second, it is social, not individual, betterment, which to us means more than greater individual utility or individual consumption. Third, we define it as incorporating greater equality in access to material, psychological, and other benefits of society, not simply a larger amount of such benefits; thus, the concept includes an explicit concern with social justice. Finally, it also allows us to recognize that societies are often better at achieving consensus about what they don't want (illiteracy) rather than what they do want (a functional definition of the minimum acceptable reading level). Social betterment provides evaluators with a *guiding star*, a term that summarizes our concern for the well-being of others and society as a whole and, we believe, estab-

lishes a clearer focus on our ultimate goals for evaluation and ourselves as evaluators. Admittedly, the specific nature of social betterment is often disputed and is iteratively sorted out over time by social and political processes. Ignoring this reality by choosing a different guiding star for evaluation does not make this reality go away, however.

Identifying social betterment as the ultimate purpose of evaluation may not seem revolutionary to thoughtful members of the evaluation community. At the same time, we believe that explicit attention to it has some noteworthy implications. First, we do not take the direct and intended use of evaluation in decision making as the ultimate criterion for judging the worth of the evaluation itself. Sometimes the impact of evaluation is indirect, delayed, and different from what an intended user might have specified. Nevertheless, such evaluation influence is important. Likewise, we suggest that evaluators and others involved in evaluation planning should not focus on maximizing the likelihood of intended use and instead should consider the risk-reward ratio of alternative approaches to evaluation. Imagine two alternative evaluations that might be conducted. Option A has a high likelihood of use, but if used would result in changes likely to have only modest if any effects on service recipients' outcomes. Option B, in contrast, has a lower likelihood of use, but if influential could make a huge difference in terms of outcomes. Taking social betterment as an ultimate goal, those planning evaluations are invited to take thoughtful consideration of these options, often tilting toward Option B. An emphasis on maximizing use sometimes can lead to a less than optimal choice among feasible evaluation approaches.

Second, in the public sector at least, the potential consumers of evaluation typically are legion, and the decision-making processes diffuse. Evaluation planning needs an approach that matches this reality. It should be more policy analytic, mindful of

the potential environment for decision making and action, rather than based on interactions with a specific, identifiable intended user.

A third implication of explicitly identifying social betterment as the guiding star for evaluation relates to methods choices. Debates in evaluation and related fields have too often focused on alleged priorities for, or frailties of, a given method such as randomized controlled trials (RCTs). Instead, the focus should be on when, how, and why alternative methods can facilitate evaluation's contribution to social betterment. Take as an example the idea of RCTs as a possible methods choice. RCTs can have relative advantage for generating unbiased estimates of the effect of a program (relative to the alternative embodied in the control or comparison group). This relative advantage is especially notable when the program is operating amidst multiple other factors that might affect the outcome(s) of interests (also known as internal validity threats) and when small or modest effects are of interest. But information about the average effect of a substance abuse prevention program, say, is most relevant when a fork in the road is ahead, for instance, when a decision will be made about which course of action, which program if any, to adopt. Thus, from the social betterment perspective, the first question involves whether the situation at hand is of this "fork in the road" variety or not. Alternatively, if no fork in the road is forthcoming but there are opportunities for evaluation to contribute to improvements of an ongoing program, an RCT or similar method of evaluating the overall program would likely not be a good choice. In addition, even if a fork in the road is anticipated, the question arises as to the relative contribution to betterment that an RCT would likely generate, in comparison to alternative methods such as various quasi-experiments, a case study, or even descriptive information from a program's set of performance indicators. As good evaluators know, and as the

idea of social betterment clarifies, method choice is not primarily about methods, but rather about the judged likelihood of a given method or combination of methods contributing to social betterment, relative to feasible alternatives, in the anticipated decision making and action context.

A fourth implication of locating social betterment as the ultimate, distal goal of evaluation is that it is necessary to identify intermediate, more proximal considerations, to guide evaluation planning. There is more than one way that evaluation can contribute to social betterment. Classic distinctions such as that between formative and summative evaluation imply as much. More generally, there are different ways that evaluation can help. Accordingly, there is not a single predominant, specific purpose for evaluations. Rather, in our view, evaluators should be mindful of a range of immediate purposes for an evaluation that could serve the ultimate goal of social betterment.

MULTIPLE PURPOSES OF EVALUATION

Our theory or model of evaluation points to alternative evaluation *purposes* as more proximal guides to evaluation planning. Early in the evaluation field, the distinction between formative and summative evaluation gave way to the idea of three general purposes of evaluation: to judge merit or worth, to improve programs, and to generate knowledge (Patton, 1997). In Mark and colleagues (2000), we expanded the list of evaluation purposes to four: overall assessment of merit and worth, when evaluation is intended to support bottom-line judgments; program and organizational improvement, when evaluation is meant to point to changes in a program or the organization providing it; accountability and oversight, when evaluation is designed to provide information about a program's compliance with legislative or other mandates; and knowledge development, when the desire is for evalua-

tion to expand understanding, for example, of the nature of a social problem or the feasibility of implementing a new approach to its solution.

Recognition that there are alternative evaluation purposes provides a starting point for evaluators to use to assess priorities. That is, a kind of situational analysis guides identification of the conditions under which each of the four evaluation purposes should rise to the top in evaluation planning. In Mark and colleagues (2000), we described a set of situations in which each purpose could rise to the top. For example, the previous discussion of whether a fork in the road is anticipated represents a major consideration in whether priority should be given to the purpose of assessing merit and worth. In contrast, when no fork in the road is anticipated but limited formative evaluation has been conducted and/or multiple variants on program practices exist, priority might instead be placed on the purpose of program and organizational improvement. Oversight and compliance might receive higher priority when failure to follow mandates (e.g., client eligibility requirements or specification of the nature of services offered) is expected to harm program effectiveness or efficiency, or more simply when this focus is demanded by those with oversight responsibilities. Knowledge development may take major priority when other contributions seem unlikely (e.g., if the evaluator comes to find the evaluation is ritualistic) or when knowledge development can coexist with or even facilitate another purpose (e.g., when study of program moderators will enhance an assessment of merit and worth). A more detailed discussion of factors that should affect the prioritization of evaluation purposes is given in Chapter 3 of Mark and colleagues (2000). No model can capture adequately all the nuances of every specific case, and considerations may change with historical shifts, such as differing legislative agendas, when attention to budget deficits changes, or with expanded

notions of evaluation purposes as the field evolves. Nevertheless, this list can assist as a guide for evaluation planning.

In general, the central idea is that evaluation planning should consider the anticipated range of potential programmatic decisions and actions that may be under consideration. The goal is to identify which evaluation purpose is likely to optimize evaluation's contribution to social betterment. Of course, a degree of uncertainty is inherent in this process. But the alternative, we believe, is even less satisfactory. And at least some of the time, there are strong enough cues. For example, initial funding for a program may be scheduled to run out a year after the evaluation is due, signaling a coming fork in the road. Or a program may be funded statewide via a voter referendum, signaling that the program will likely continue for the foreseeable future, but with the door open to improving the target population's outcomes by determining which patterns of implementation are associated with better outcomes.

With an evaluation purpose (or a sequence or combination of purposes) in mind, choosing from among methods can be done more sensibly. Evaluation purpose should lead to evaluation questions, and these questions should limit the range of viable methods. For example, the purpose of program improvement often suggests three questions: Does program implementation vary? Do program outcomes vary? Are there variations in implementation that are systematically associated with variations in outcomes? In this case, comparative case studies and analysis of descriptive information from administrative and/or survey data seem to be likely methods choices. This in turn leads to additional questions, such as "How do we choose cases to maximize the ability to observe implementation variation?" or "How do we sample clients to provide the most accurate description of outcomes?" Such questions are answered iteratively, injecting the issue of feasibility into the process whenever an answer is ten-

tatively agreed upon. Method choices, we believe, are best made in service of a deliberately chosen purpose that represents a pathway to social betterment, by addressing a series of questions that involve trade-offs between the ideal and the possible and making the choices transparent for consumers of the information produced by the evaluation.

The path between the planning and implementation of an evaluation and its potential contributions to social betterment runs through evaluation influence. Although the term *evaluation influence* was intended to capture all the consequences of evaluations, our focus is primarily on those that can lead to social betterment, whether through the adoption of a new policy or the continuation, termination, or reform of an existing program. In the next section, we discuss both the distinct contributions that stem from focusing on *evaluation influence* rather than previously used terms, and from including attention to the processes through which evaluation influence may occur.

ALTERNATIVE PATHWAYS TO EVALUATION INFLUENCE

The literature has long differentiated between direct, conceptual, process, and symbolic use, and somewhat more recently, process use (Caracelli, 2000; Cousins & Leithwood, 1986). However, since the turn of the century, several evaluation theorists have argued that "evaluation influence" rather than "evaluation use" should be the focus (Henry, 2000; Kirkhart, 2000; Mark & Henry, 2004). In addition to the direct consequences intended by evaluators, sponsors of evaluation, and other stakeholders, influence includes those consequences of evaluation that are indirect, unintended, and distant from the evaluation itself, perhaps even without the relevant party's awareness of the influence. For example, a legislator's vote for a new program may be indirectly influenced by evaluation findings of the

program's effectiveness in another state. The findings may have motivated the persuasive efforts of an aide, even though the legislator is not aware the evaluation exists. Although we endorse a focus on evaluation influence, we are most concerned about trying to bring about intended uses and avoid unintended negative consequences. We also encourage making explicit choices about how an evaluation should aim to achieve its intended consequences.

In a pair of papers (Henry & Mark, 2003; Mark & Henry, 2004) and an extension in an American Evaluation Association address (Mark, 2006), we have attempted to identify different categories of consequences that evaluation may have, as well as various processes that might help lead to these consequences. While the full-blown representation of these consequences is rather complicated, we can summarize some of the highlights to provide a perspective on the bigger picture features of our model of evaluation influence.

One feature of the model is that for each type of evaluation consequence, it distinguishes among three "levels of analysis": individual, interpersonal, and collective. The basic idea is that the consequences of evaluation sometimes involve a change within a particular person (individual), sometimes a change in the interaction between individuals (interpersonal), and sometimes a process at a more macro, organizational unit (collective). For example, consider instances in which evaluation findings suggest a behavioral change, that is, an action of the sort that could be classified under the older label of direct or instrumental use. As our model suggests, the desired behavioral change might be at any of the three levels. For example, in an educational evaluation, the behavioral change might involve individual classroom teachers' instructional practices or, alternatively, the initiation of certain collaborative practices across teachers (interpersonal) in a school. Or, instead, the collective-level adoption of a program policy or program by a legislature.

The distinction among these different levels of analysis is not trivial. As Henry and Mark (2003) suggest, the pathways that lead to individual or interpersonal change may be quite different than the pathways that lead to collective action (also see Weiss, 1998, pp. 264–265). By getting beyond relatively gross categories such as "direct use," evaluators might describe a more complete set of pathways and incorporate them into an evaluation influence plan to increase the likelihood that evaluation influences action and moves toward social betterment. Explicit recognition of the different levels of evaluation's influence can also help clarify debates and differences within the field. For example, we believe that advocates of RCTs and meta-analyses implicitly focus more on collective change, such as program funding decisions, while evaluators interested in lived experience and practitioner judgment may focus more on the individual (or perhaps interpersonal) level.

In addition to differentiating the three levels of analysis, Mark and Henry (2004) identify four categories of evaluation consequences. Two are familiar. "Behavioral processes," according to Mark and Henry, refer to changes in actions. This category thus includes instrumental use but is broader, including for example influence the person is not aware of, such as changes in response to social norms that an evaluation influenced. "Cognitive and affective processes" refer to shifts in thoughts and feelings, such as attitude valence. This corresponds to conceptual use but, again, is broader. In contrast, two categories, general influence and motivational processes, are typically of interest not as the ultimate change that an evaluation might stimulate, but as the initial or interim steps along a pathway to that change. "General influence processes," according to Mark and Henry, are not likely to be the intended end of an influence process but "are likely to set into motion some [other] change." For example, the individual-level process of "elaboration" refers to a person thinking more deeply

about an issue—a process that extensive research shows can trigger long-lasting change in attitudes and perhaps behaviors. Similarly, Mark and Henry describe “motivational processes” as involving goals, aspirations, and responses to perceived rewards and punishments. Such motivational processes may play an important role in changing behavior, as when highly visible and recognizable dissemination of brightly colored bins for recyclable materials lead homeowners to fear social disapproval if they do not participate in a curbside recycling program.

Mark (2006) proposed the addition of another category of evaluation consequences to the Mark and Henry (2004) framework. He labeled this new category as “relational consequences” because it includes efforts by evaluators to modify not behavior or attitude but aspects of ongoing relationships, structures, and organizational processes. The relational category includes, for example, such potential consequences as individuals’ self-perception of their empowerment (Fetterman, 1996), creation of a democratic forum for deliberation (House & Howe, 1999), and the facilitation of the learning organization (Preskill & Torres, 1998). In general, the evaluation theories (or models) that advocate for such relational consequences are more recent, relative to earlier theories that focus on program-related action and understandings (Alkin, 2004). And they often tend to see evaluation not as narrowly concerned with the policy, program, or practice being evaluated; rather, these approaches often treat evaluation in part as a vehicle to use to try to bring about other forms of change, such as empowerment or the creation of a learning organization. As a result, many of these approaches may be subject to the criticism, “But is it evaluation?” from commentators who hold a more traditional definition of evaluation (e.g., Scriven, 1997). They are also subject to questions about whether in practice they achieve their stated relational goal; for example, Miller and Campbell (2006) ask whether evidence indicates that

empowerment evaluation actually empowers people. However, these criticisms are not unlike ones that can be applied to traditional views about the intended consequences of evaluation, such as whether classic comparative evaluation designs actually are used in decision making. Although not established in the literature, the relational category or something like it seems useful, in part because it reminds us that there is a set of potential evaluation consequences that extend beyond the boundaries of the traditional direct and conceptual use.

Here, we have only sketched the general structure of the influence process model from Henry and Mark (2003), Mark and Henry (2004), and Mark (2006). Admittedly, the model is complex. But that is appropriate, we believe, because the pathways by which evaluation becomes influential are complex and varied. In a sense, the model reflects our belief that evaluators benefit by considering evaluations themselves as interventions and by adopting something akin to program theories for evaluations. From this vantage, the influence model is like a catalog, a taxonomy, or a listing of a large array of processes and consequences, only a subset of which might operate for any particular evaluation.

Again, we believe there are nontrivial implications. First, the model invites evaluators to try to draw out the expected pathway by which an evaluation is expected to be influential. Our sense is that the traditional “end states” of instrumental and conceptual use do not capture the potentially long and complex path through which evaluation often comes to have its influence. Consider a simplified example. A briefing about evaluation findings might trigger (1) a legislative aide in attendance to engage in elaboration of her thoughts about a program, (2) which in turn leads her to have a more positive attitude toward the program, (3) which then leads her to engage in interpersonal processes such as persuasion efforts directed at other legislative staff and her boss, (4) which might eventually contribute to the inclusion

of language that would expand program funding in an appropriations bill, leading to (5) a collective-level policy change. The reality is likely to be quite complex, with some changes predictable and others emergent as the process develops, and even pushback in directions other than those anticipated by the evaluator.

Second, we believe there is considerable promise for research and practice if increased attention were given to these kinds of “influence pathways.” Thinking about influence pathways could be a useful guide to new kinds of research on evaluation influence and use. For example, case studies could trace the influence pathways from an evaluation (e.g., Herbert, 2022). Experiments could examine the effects of arrangements designed to facilitate (or inhibit) a process of potential interest, such as elaboration (e.g., Rykaczewska, 2021). In terms of practice, if evaluators seriously consider possible influence pathways during evaluation planning, they might take steps to increase the likelihood that evaluation will matter.

Thinking about influence pathways could also sharpen our thinking about alternative evaluator roles in facilitating use and influence. In our view, there has long been a kind of dichotomous thinking about the role of evaluators, either suggesting the evaluator’s primary responsibility is to facilitate use (e.g., Patton, 1997), or suggesting more of a hands-off role of provider of information (e.g., Campbell, 1973). Attention to alternative roles has perhaps been inhibited because there has not been a common vocabulary to identify the multiple steps in an influence pathway. And different potential roles exist. For example, an evaluator working for an advocacy group or private foundation might stay involved all along the multiple series of steps in an influence pathway. An evaluator working for an independent government agency might be involved only for a few steps, perhaps relying on advocates, lobbyists, or other actors in the policy process to move beyond the initial steps.

Evaluators at universities or research firms might operate in various ways, depending on their personal preferences or those of their sponsors, with one approach involving determined dissemination to others in the policy and practice communities related to the program while not taking responsibility for achieving the downstream portion of the influence pathway (Greenberg, Mandell, & Onstott, 2000).

EPISTEMOLOGICAL ROOTS

The editors asked us to address our epistemological roots. In Mark and colleagues (2000), we claimed allegiance to a form of realism. Our realism is something close to the “critical realism” of Campbell, which assumes that one’s access to the world is imperfect, mediated, and subject to bias. In part, we endorsed realism to differentiate our approach from what we see as anti-realist claims in the rhetoric of some evaluation theorists. As Bhaskar (2008, p. 33) said, “To be a fallibilist about knowledge, it is necessary to be a realist about things.” At the same time, we recognize the importance of human and social constructions. These apply to evaluation planning, such as in the choice of values that inform evaluation design and measures, as well as to the perceptions and actions of program clients and others. Nevertheless, recognition of social construction does not, we think, imply that efforts to estimate program effectiveness are inherently somehow wrong. Nor does it obviate such notions as bias in inference, construct and external validity, or gaming a testing system.

There is a way in which, in retrospect, we wish there was no press for evaluators to lay claim to an epistemological position. Too often, these claims are associated with excess, at least in rhetoric if not in evaluation practice. Perhaps evaluation theorists need to follow the lead of most humans, including sensible evaluation practitioners, who in everyday life exhibit a combination

of realism (e.g., stopping at a red light) and constructivism (e.g., recognizing that most traffic laws are arbitrary social constructions and that people have differing views about risk taking). Moreover, most people realize that effective action does not require investing heavily in determining the relative priority of these or other philosophy of science positions.

CONCLUSION

Evaluation can be directed toward several alternative purposes. Evaluation can also be influential in various ways, with even more varied pathways to influence. Indeed, most of the chapters in this book suggest one or another possible consequence for evaluation, along with recommendations about the best way of getting there. This brings to mind the question: Under what conditions should each potential consequence of evaluation take priority? Which calls forward two other questions: By what process should one judge the relative priority of these alternatives? And, given priority to a particular kind of evaluation consequence, how should one try to achieve it? Attempting to answer these questions is at the core of the work we have done to contribute to the development of evaluation theory.

Allegiance to the idea of evaluation as assisted sensemaking implies a kind of modesty. Humans do not have access to what philosophers occasionally call the God's-eye view. Nor do evaluators have access to a God's-eye view of how evaluation should be done. Our own ideas about evaluation have evolved over time, and we anticipate they will continue to evolve through processes akin to mutation, adaptation, and selection. We look forward to further influences by colleagues on and off the evaluation theory tree. And we heartily root for their efforts and for those of all who seek to contribute, in one way or another, to the endeavor of social betterment through evaluation.

REFERENCES

- Alkin, M. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences*. SAGE.
- Bhaskar, R. (2008). *A realist theory of science* (2nd ed.). Routledge.
- Campbell, D. T. (1973). The social scientist as methodological servant of the experimenting society. *Policy Studies Journal*, 2(1), 72.
- Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. In V. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use* (New Directions for Evaluation, Vol. 88, pp. 99–111). Jossey-Bass.
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization, *Review of Educational Research*, 56(3), 331–364.
- Fetterman, D. M. (1996). *Foundations of empowerment evaluation*. SAGE.
- Greenberg, D., Mandell, M., & Onstott, M. (2000). The dissemination and utilization of welfare-to-work experiments in state policy-making. *Journal of Policy Analysis and Management*, 19, 367–382.
- Henry, G. T. (2000). "Why not use?" In V. Caracelli & H. Preskill (Eds.), *Evaluation use* (New Directions for Evaluation, Vol. 88, pp. 85–98). Jossey-Bass.
- Henry, G. T. (2002). Choosing criteria to judge program success: A values inquiry. *Evaluation*, 8(2), 182–204.
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24, 294–314.
- Henry, G. T., Pham, L., Kho, A., & Zimmer, R. (2020). Peeking into the black box of school turnaround: A formal test of mediators and suppressors. *Educational Evaluation and Policy Analysis*, 42(2), 232–256.
- Herbert, J. L. (2022). A study of evaluation influence in two child protection programs. *Evaluation Review*, 46(6), 683–708.
- House, E., & Howe, K. (1999). *Values in evaluation and social research*. SAGE.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000, 5–23.
- Mark, M. M. (2006). *The consequences of evaluation: Theory, research, and practice*. Presidential address at the annual meeting of the American Evaluation Association. Portland, OR.

- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*(1), 35–57.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. Jossey-Bass.
- Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation, 27*, 296–319.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). SAGE.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). SAGE.
- Preskill, H., & Torres, R. (1998). *Evaluative inquiry for organizational learning*. SAGE.
- Rykaczewska, A. M. H. (2021). *Achieving evaluation influence through elaboration likelihood model-informed evaluation product designs*. CGU Theses & Dissertations No. 225. https://scholarship.claremont.edu/cgu_etd/225.
- Scriven, M. (1997). Empowerment evaluation revisited. *Evaluation Practice, 18*, 165–175.
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice, 9*(1), 5–20.
- Weiss, C. H. (1998). Improving the use of evaluations: Whose job is it anyway? *Advances in Educational Productivity, 7*, 263–276.
- Zimmer, R., Henry, G. T., & Kho, A. (2017). The effects of school turnaround in Tennessee's achievement school district and innovation zones. *Educational Evaluation and Policy Analysis, 39*(4), 670–696.

CHAPTER 6

Theory-Driven Evaluation Approaches Conventional and Innovative

Huey T. Chen

INTRODUCTION

Theory-driven evaluation (TDE) goes beyond assessing whether a program works or does not work by empirically illustrating how and why. TDE also provides valuable and insightful information for stakeholders to better understand and improve their programs. Initially, TDE was introduced by Chen and Rossi (1980) and Chen (1990) as a contrast to method-driven evaluation that was popular during that period. Since then, TDE has been widely applied by evaluators and has become one of the major evaluation perspectives (Chen, 2005, 2015; Coryn, Noakes, Westine, & Scholer, 2011; Donaldson, 2007). Under TDE, some evaluators have used different labels for this type of evaluation, including theory-based evaluation (Rogers & Weiss, 2007), theory of change (Fulbright-Anderson, Kubisch, & Connell, 1998), realistic evaluations (Pawson & Tilley, 2007), and process tracing (Checkel, 2008). Despite the different labels, they all empha-

size that evaluation is not an atheoretical undertaking, and stress the importance and usefulness of program theory in evaluation.

This chapter provides a guide on developing program theory and introduces various types of TDE approaches available for practitioners. The chapter is organized in four sections that focus on (1) how to facilitate stakeholders in developing a program theory to improve communications and coordination of program activities, (2) discussing conventional TDE approaches, (3) introducing innovative TDE approaches, and (4) how to use TDE to investigate vital but understudied phenomena.

A GUIDE TO FACILITATE STAKEHOLDERS' CLARIFICATION OF THEIR PROGRAM THEORY OR THEORY OF CHANGE

When stakeholders design and develop an intervention program, they implicitly make two sets of assumptions: *prescriptive*, outlin-

ing actions necessary to initiate and support the desired change processes, and *descriptive*, explaining the causal mechanisms linking the intervention to desirable outcomes (goals). For an intervention program to be effective, both prescriptive and descriptive assumptions must work well (Chen, 1990, 2005, 2015). A configuration of both prescriptive and descriptive assumptions is called *program theory* or *theory of change*. A program theory is useful for stakeholders to improve communication and coordination of their program. Despite the importance of having a clear understanding of program theory, stakeholders hardly make them explicit in program documents. TDE stresses the evaluators' assistance is in helping stakeholders clarify their program theory is essential. A program theory not only can help stakeholders manage their program but it also serves as a foundation for evaluators to conduct a TDE.

The facilitation process requires using a model to guide the discussion with stakeholders. Popular models include logic models, causal mechanisms, and causal loop diagrams. I recommend the action model-change model schema (Chen, 2015) for the following reasons: (1) The schema addresses both prescriptive and descriptive assumptions, (2) the schema can be applied to intervention programs with different levels of complexity, and (3) the schema provides a foundation for developing different TDE approaches that are discussed below.

The Action Model–Change Model Schema

The schema consists of two related models: the action model and the change model. The *change model* describes the causal process generated by the program. It makes the descriptive or causative assumptions underlying the program explicit through the following three elements:

1. *Goals or outcomes*, which reflect the desire to fulfill unmet needs, such as poor health, inadequate education, or poverty.

Outcomes are the concrete, measurable aspects of the goals.

2. *Intervention or treatment*, which includes all program activities that aim directly at attaining goals or outcomes. In other words, the intervention is the agent(s) of change within the program.

3. *Causal mechanisms*, which mediate between the linkages between the intervention and outcomes. A program must activate the mechanisms first in order to reach goals. Causal mechanisms have different names, including mediators, intervening variables, and leverages.

The *action model* is a systematic plan for arranging staff, resources, settings, and support organizations to reach a target group and deliver intervention services. It makes the prescriptive assumptions of the program explicit. The action model consists of the following elements:

1. *Implementing Organization: Assess, Enhance, and Ensure Its Capabilities*. The program relies on the organization to allocate resources and coordinate activities. The organization must recruit, train, and supervise implementers and other staff. It is crucial to ensure that the implementing organization can perform these functions.

2. *Program Implementers: Recruit, Train, and Maintain Competency and Commitment*. Program implementers are responsible for delivering services to clients and include counselors, case managers, outreach workers, school teachers, health experts, social workers, and so forth. The implementers' qualifications, competency, commitment, enthusiasm, and other attributes can directly affect the quality of service delivery and outcomes.

3. *Associated Organizations/Community Partners: Establish Collaborations*. Programs often benefit from, or even require, cooperation and/or collaboration between the implementing organization and other

groups or organizations. If linkage or partnership with these organizations is not established properly, implementation of the program may be hindered.

4. *Intervention and Service Delivery Protocols.* The *intervention protocol* is a curriculum or prospectus stating the exact nature, content, and activities of an intervention—in other words, the details of its orienting perspective and its operating procedures. The *service delivery protocol* refers to the particular means such as face-to-face meetings, group meetings, or phone calls to take to deliver the intervention to the target population group.

5. *Ecological Context: Seek Its Support.* Some programs have a particular need for *contextual support*, which means that they need a supportive environment for the program to work. Both *microlevel contextual* and *macrolevel contextual support* can be crucial to a program's success. Microlevel contextual support includes social, psychological, and material supports from families or friends that clients need to ensure their continued participation in the intervention program. In addition, the *macrolevel context* of a program, including community norms, cultures, and political and economic processes, can facilitate a program's success.

6. *Target Group: Identity, Recruit, Screen, Serve.* The target group represents the population group that the program aims to engage. Crucial assumptions about the target group that must be met include valid eligibility criteria for participation in the program; the feasibility of reaching and effectively serving the target group; and the willingness of potential clients to become committed to or at least agreeable to joining the program.

The components of the two models and their relationship are illustrated in Figure 6.1, which shows that the action model must be implemented appropriately to activate the “transformation” process in the change model. For a program to reach goals, its

action model must be sound and its change model plausible; its implementation also must be effective. This conceptual framework of program theory should be useful when designing the evaluation of the program; it is expected to produce information that will clarify the dynamic processes that lead to program success.

Guide for Applying the Schema to Help Stakeholders Clarify Their Program Theory

Funding agencies usually ask stakeholders to implement an evidence-based or homegrown intervention program. Since an evidence-based intervention is based on a behavioral/social theory, the theory provides sufficient information for developing a change model for the program. In this case, evaluators need to confirm the theory and change model with stakeholders and focus on clarifying their action model. However, since a homegrown program is not based on a formal behavioral/social theory, evaluators must facilitate or assist them in clarifying their action model and change model. Evaluators can apply either face-to-face interviews or focus groups for these purposes. For small programs, face-to-face interviews are recommended. Focus groups work better for medium or large programs. The evaluators must also apply one or more of the following approaches in facilitating stakeholders clarifying their action model and change model.

The first theorizing approach, *backward reasoning*, starts with the change model and moves back, toward the action model, to obtain the arguments that justify the program. The reasoning moves in the opposite direction to the sequences shown in Figure 6.1. Backward reasoning starts with the question: *What goals the program seek to achieve?* It then continues to inquire: *Which mechanisms of these goals should the program focus on?* and *What intervention will affect these mechanisms?* When the change model has been completed, evaluators can facilitate stakeholders' development of the cor-

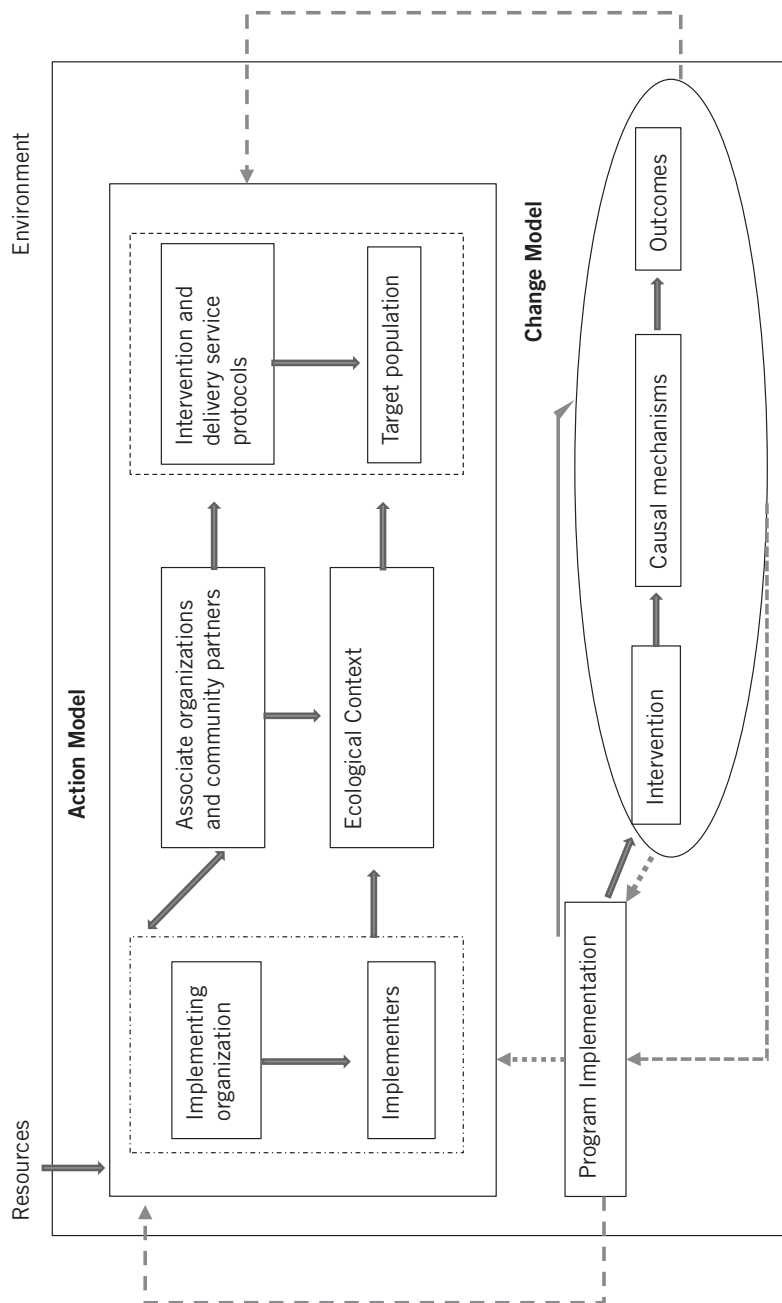


FIGURE 6.1 Action model–change model schema.

responding action model. This can be achieved by clarifying the following: *which target groups need to be reached and served, what kind of program implementers and implementing organizations will suit the program, what types of intervention and implementation protocols seem best, what partners are suited for collaboration to achieve program goals, what environmental support is necessary and how can it be achieved.*

The second theorizing approach, *forward reasoning*, follows the flow outlined in Figure 6.1, the action model first, then the change model. Forward reasoning starts by considering the capacity of the organization to excel in certain areas and continues to identify components that complement that capacity to ensure program success. Forward reasoning should clarify the following: *which intervention and implementation protocols the implementing organizations excel at, what group needs to be reached and how, what setting and delivery protocols make sense, whether clients face barriers to receiving services, and how the program can alleviate them, and how and where contextual support should be sought for the intervention.* When the action model is complete, evaluators and stakeholders can develop a change model by asking two questions, in sequence: *What mechanisms will be changed by the intervention? What outcomes will be achieved by activating these mechanisms?*

The third theorizing approach, *forward-backward reasoning*, is possible since the two are not mutually exclusive. They can be used back and forth, to facilitate stakeholders' discussion to make explicit their action model-change model schema. Forward-backward reasoning is more time-consuming than the other two approaches, but it may bring to bear the best of both approaches (Chen, 2015).

Examples of Assisting Stakeholders to Clarify Their Program Theory

The following two examples reflect the experience of myself and that of my team by applying the schema to assist stakeholders in clarifying their program theories.

Example 1: Retention Program

A homegrown retention program was created to assist disadvantaged nursing students to be successful in school and graduate (Chen, Morosanu, Bellury, & Kimble, 2019). During a working group meeting, the evaluation team used the schema and the theorized reasoning approaches previously described to assist stakeholders in clarifying major components of the program. The resulting schema is illustrated in Figure 6.2.

Stakeholders found the process and the resulting schema very useful for coordinating program activities. The evaluation team used the schema as a basis to develop a multiwave formative evaluation that is discussed later in this chapter.

Example 2: Learning Community

Educators of Taiwan launched the Learning Community Program in 2013. During implementation, stakeholders realized they had difficulties describing and communicating information about the program due to its intricate and dynamic interactions. Evaluators used the schema to help stakeholders clarify the program. The Learning Community Program was complex, with multiple layered interventions. The evaluation team expanded the schema into three change models for the school, teacher, and student levels, to synthesize the program (Chen, Pan, Morosanu, & Turner, 2018).

The resulting schema is presented in Figure 6.3.

CONVENTIONAL THEORY-DRIVEN EVALUATION APPROACHES

There are two types of conventional TDE approaches: outcome and process.

Theory-Driven Outcome Evaluation

Theory-driven outcome evaluation is to assess the entire change model rather than

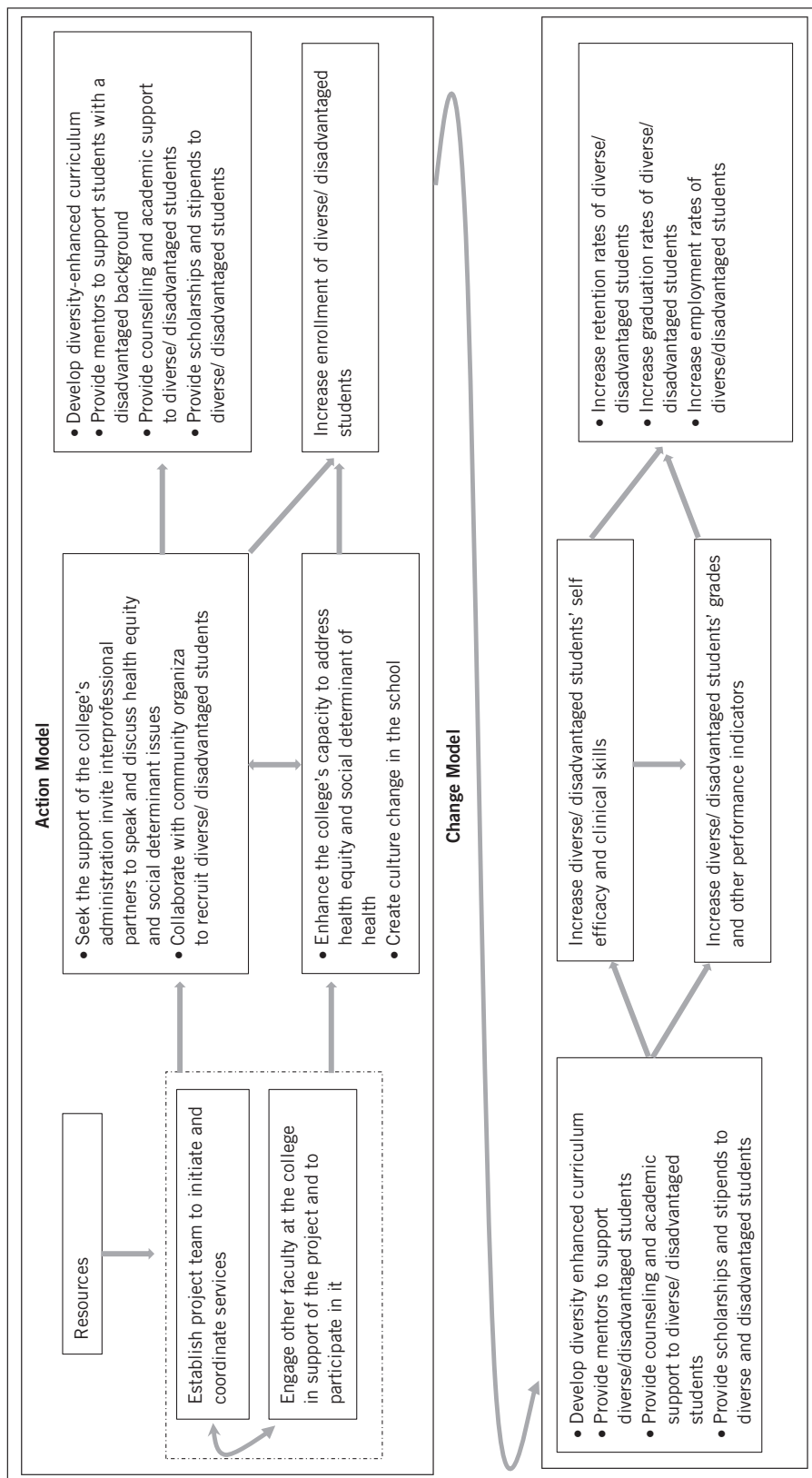


FIGURE 6.2 The action model–change model schema of the Nursing Student Retention Program.

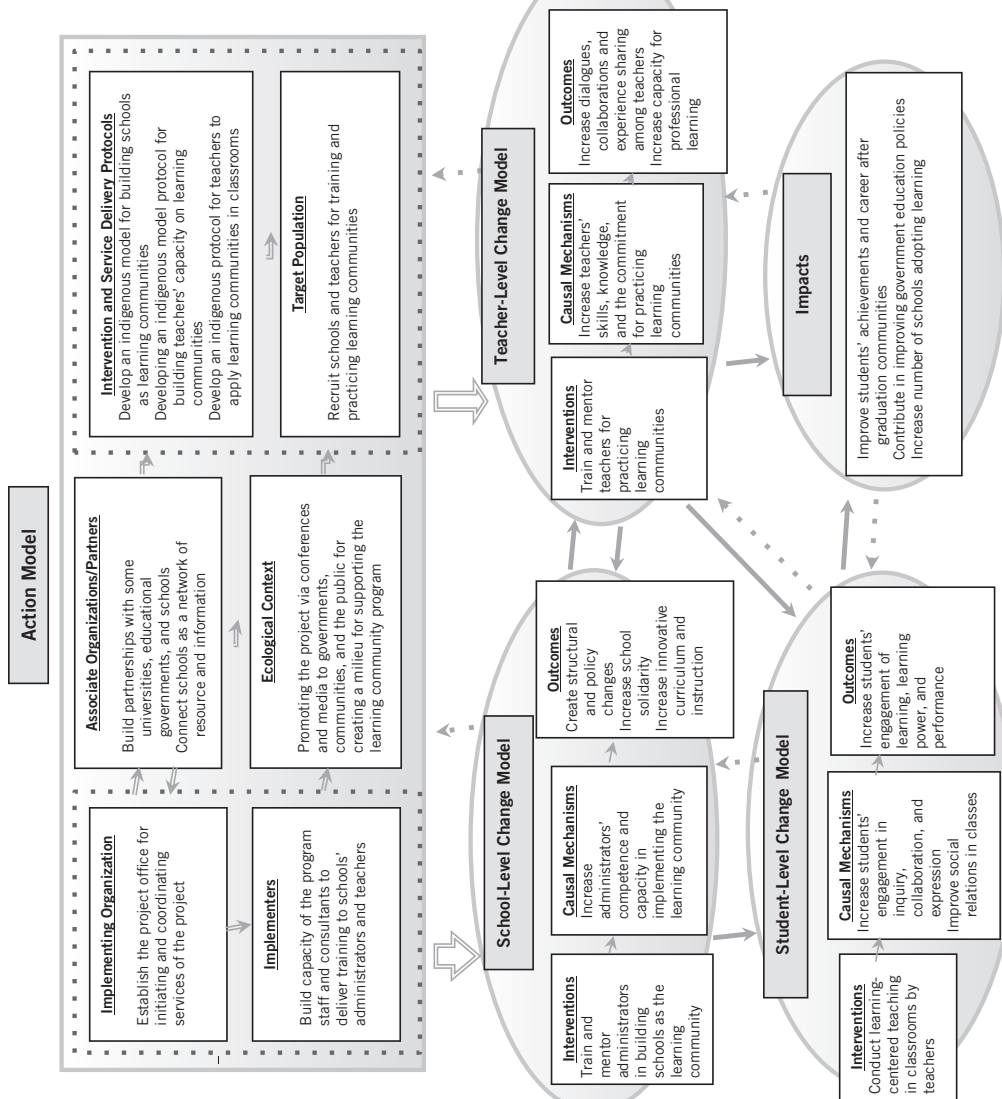


FIGURE 6.3 Action model—change model of the Learning Community Program.

the relation between an intervention and outcomes. It consists of two steps: The first is to assess whether the intervention affects the mechanism; the second, whether the mechanism affects the outcomes. This kind of evaluation provides information not only on whether an intervention has an impact on outcome variables, but also why. Using the retention program discussed previously as an example, a theory-driven outcome evaluation would assess whether the program increases disadvantaged students' self-efficacy, clinic skills, and grades; and then assess whether these mechanisms increase students' retention and graduation rates as shown in Figure 6.2. Theory-driven outcome evaluation has been the most widely used TDE approach. Chen (2015), Donaldson (2007), and Rogers and Weiss (2007) provide concrete examples of this application.

Theory-Driven Process Evaluation

Process evaluation usually refers to the assessment of whether an intervention is implemented with fidelity or adaptation.

TDE argues that process evaluation focusing only on intervention fidelity tends to be too narrow because it ignores contextual factors that influence the implementation process (Chen, 2015). TDE proposes a theory-driven process evaluation approach to address the problem. Theory-driven process evaluation is to assess the implementation of the entire action model. A good example to illustrate theory-driven process evaluation is an assessment of the implementation of a school-based anti-drug abuse program for middle school students in Taiwan (Chen, 1997). The Ministry of Education developed the program and required teachers to identify students with drug abuse problems and provide them with counseling services. Evaluators had a meeting with representatives of the ministry, and another meeting with school representatives, to facilitate clarification of the action model of the program. The participants in the two meetings agreed on the major components of the program as illustrated in the second row of Table 6.1.

Evaluators used mixed methods to collect data on how the action model was imple-

TABLE 6.1. The Spring Sun Program: Normative versus Actual

Program domains/dimensions	Normative	Actual
Goal/outcome	Reduction of student drug use to be verified through urinalysis	Reduction of drug use, but urinalysis collection environment not controlled
Treatment	- Primary: provide quality counseling to abusers - Secondary: basic drug education	- Primary: counseling mainly involved use of threats, admonishment, and/or encouragement not to use - Secondary: basic drug education
Implementation environment		
Target group	All drug-abusing students	Only those drug abusing students who were easy to reach
Implementors	Teachers provided with adequate drug treatment training and information	Teachers lacked adequate drug treatment skills and information
Mode of delivery	Compulsory individual counseling	Compulsory individual counseling; but with problems such as lack of plan, format, and objective
Implementing organization	All schools that can adequately implement the program	Smaller schools had difficulty implementing the program
Interorganizational procedures	Effective centralized school system	Communication gap, mistrust between Ministry of Education and the schools
Micro-context	Eliminate video game arcades	Video game arcades still exist
Macro-context	Strong public support	Strong public support, but problematic education system (elitism)

mented. The results are presented in the third row of Table 6.1. The schools severely modified the planned components to fit their capacity, as indicated in Table 6.1. The discrepancies between the planned and implemented action model were due to a lack of appropriate counseling training, the overburdening of teachers with counseling work with no change to their usual teaching responsibilities, and a lack of communication and mistrust between an authoritarian ministry and the school system.

INNOVATIVE THEORY-DRIVEN EVALUATION APPROACHES

The advancement of program evaluation requires the development of innovative approaches. This section is an overview of TDE's contributions on this front.

Proactive Evaluation

Traditional evaluation approaches tend to focus on assessing the past performances or merits of a program's implementation or goal attainment. However, since stakeholders run the program on a day-to-day basis, they are also very interested in proactively gathering information that can help them effectively implement the program. Currently, evaluators have mainly relied on formative evaluation for this purpose. TDE, however, can offer more.

Multiwave Formative Evaluation

More and more agencies ask grantees to improve their programs on an ongoing basis. While this is a requirement for funding, few grantees are conducting an ongoing assessment of their implementation process; the common practice is the one-time formative evaluation during any given implementation cycle. The multiwave formative evaluation approach was proposed in response to the rapid-cycle improvement requirement. Multiple iterations of the tra-

ditional formative evaluation allow evaluators to assess problems and barriers of the program, and to inform the stakeholders' ongoing troubleshooting process. Each subsequent wave is a reassessment that eventually leads to "all creases being ironed." The action model-change model schema is useful in providing the necessary conceptual framework for this purpose. The multiwave formative evaluation was pioneered to evaluate a retention program for disadvantaged nursing students (Chen, Morosanu, Bellury, et al., 2019). This particular evaluation comprised three waves of formative evaluation. Each wave employed face-to-face interviews with several students (the target population receiving services) and program implementers. The management team used the results of each evaluation wave to improve the implementation deficits. An important lesson learned from this study is that any program is likely to have different problems or barriers *at different times*. A one-time formative evaluation can only detect these problems at a particular time but is sure to miss other problems that may arise thereafter. For example, the problem of racial tension did not surface during the first wave, but did during the second and third waves of evaluation.

Program Stress Test

Formative evaluation or multiwave formative evaluation can detect immediate problems or barriers during the implementation process. However, they are unlikely to identify the underlying functional or structural problems of the program. The program stress test was developed for this purpose. The stress test is a holistic assessment of a program, through a collective brainstorming exercise aiming to identify weak structural components and functional links unlikely to withstand implementation under real-world conditions (Chen et al., 2018). Such exercise requires thorough knowledge of the program; therefore, the participants should be key stakeholders with

such knowledge and experience in managing, coordinating activities, and implementing the program. Evaluators should serve as facilitators to maintain the focus of the conversation and ask relevant and sometimes delicate questions that can unearth potential weaknesses of the program. The ex-post program stress test was used to diagnose the retention program mentioned earlier. The management team had used the results to improve the program.

Holistic Effectuality Evaluation

For several decades, evaluators applied the concepts and methodology provided by the Campbellian validity typology (Shadish, Cook, & Campbell, 2002) to assess social and behavioral programs. The Campbellian typology provides the theory and methods for rigorously assessing whether the outcomes can be attributed exclusively to an intervention (internal validity). Therefore, it aims to identify the *pure and independent effects* of the intervention. While TDE recognizes the scientific merits of the Campbellian typology and its contributions, it nevertheless argues that this typology should be used as an option rather than as the “gold standard” for outcome evaluation. TDE stresses that stakeholders use interventions to serve the clients’ needs rather than prioritize a research design. Furthermore, stakeholders often need to *contextualize the intervention*, to precisely fit the particularities of the community. Accordingly, the intervention impacts are the result of both the intervention and contextual factors rather than the intervention alone. For this reason, the Campbellian typology might not be capable of catching the real-world effects of interest to stakeholders. To address the problem, TDE provides the holistic effectuality evaluation approach as an alternative for assessing the real-world effects of interventions (Chen, 2015).

The threats to internal validity mentioned by Campbell have the potential to function as *adjuvants* to the intervention; after all,

stakeholders are interested in capitalizing on everything that can help their clients. It would be counterproductive to ignore factors that can improve the chances of intervention success in the real world. The holistic approach provides the methodology for assessing the real-world effects of an intervention (Chen, 2015).

Culturally Responsive Theory-Driven Evaluation

The culture and other contextual factors can influence how an intervention program serves clients and its impacts in a community. Culturally responsive evaluation as an evaluative inquiry that addresses the cultural context of the community and a strategy for advocating for underserved and underrepresented groups and communities (Hood et al., 2014). Bledsoe and Donaldson (2015) argued that TDE is a valid approach in articulating the situations, contexts, and circumstances of communities needed to conduct a culturally responsive evaluation. For example, in an evaluation prenatal service program for young African American mothers, TDE was useful for the identification of moderators, such as cultural context and economic history, in discerning how the program would work, under what conditions, and how potential would resonate with the intervention efforts.

Exhibited Generalization Approach

The Campbellian validity typology also contributes to providing the concept of external validity for evaluators to consider generalizability issues in evaluation. External validity refers to whether causal relationships can be generalized to different measures, persons, settings, and times (Shadish et al., 2002). However, TDE argues that while this definition is suitable for research, it doesn’t serve evaluation well given that evaluation is constrained by time and resources, and the impossibility to conduct open-ended quests of generalizability. TDE proposes a realistic and practical approach called *exhib-*

ited generalization, to address generalizability or transferability issues (Chen et al., 2014). Exhibited generalization refers to provision of sufficient information on the contextual factors supporting the effectiveness of an intervention for informing potential stakeholders or users to consider whether or not to apply the intervention for their community. By matching contextual factors, stakeholders or potential users can make an informed decision on whether the intervention program would be transferable to their community. Chen and colleagues illustrated how to apply the exhibited generalization approach in an evaluation of a carbon monoxide alarm ordinance.

USING THEORY-DRIVEN EVALUATION TO INVESTIGATE VITAL BUT UNDERSTUDIED ISSUES

Assessing Human Factors in Interorganization Collaborations

Many programs or projects require inter-organizational collaboration because these collaborations can save money and expand services. However, human factors are rarely included in the assessment across literature. The application of TDE in assessing inter-organizational projects provides an opportunity for understanding potential strains and tensions between organizations and strategies on how to address the problems. These tensions arise because different organizations have different interests and views about the program. A program's quality and effectiveness may be diminished or damaged if different interests and views about the program are not addressed. This issue was illustrated in assessing the implementation of a healthy community's monitoring and evaluation system (Chen, Morosanu, Powell-Threets, Lian, & Turner, 2019). The state health department provided a grant, guide, and training for health districts to implement the program and report quarterly and annually to the monitoring and evaluation (M&E) system. Many health dis-

tricts did not meet this requirement. Evaluators also found that the M&E system did not adequately capture the actual work of the districts in communities as related to the work plan. This study identified several factors that negatively affected the performances of the M&E. Some of the factors were that the state and health districts did not have a shared vision on how to develop and implement the program, or what to report to the M&E system. As a result, the quality of information provided by the M&E was low. The study proposed strategies that could improve the timeliness and content of reports to the M&E, as well as its utilization in the future.

Assessing Relative Strengths and Limitations of Using Behavioral Theory versus Stakeholder Theory to Develop Intervention Programs

The source of a program's change model in a schema could be a behavioral theory or stakeholder theory. An interesting and important issue is which theory is preferable. Researchers are likely to argue that behavioral theories are superior to stakeholder theories because behavioral theories have been extensively studied, which is not the case for stakeholder theory. However, only one empirical study (Chen & Turner, 2012) has formally examined this assumption. This empirical study assessed the relative effectiveness of the health advisor model (based on social learning theory) compared to a signage intervention (based on a stakeholder theory) in reducing secondhand smoking in a housing complex. The study found that the signage intervention reduced secondhand smoking while the health advisory model (CHA) did not. The data indicated that self-efficacy reduced the targeted outcome, but CHA failed to affect self-efficacy. The findings suggest that self-efficacy could affect outcomes; however, it would be challenging to change participants' self-efficacy in community settings such as a housing complex. On the other hand, the signage intervention based on the

stakeholders' theory was easy for residents to understand and follow. More studies on the relative strengths and limitations of behavioral theory versus stakeholder theory would have the benefits of cumulating not only researchers' scientific knowledge but also stakeholders' experiences and skills in working with communities for developing more realistic, effective, and sustainable interventions in the future.

Proposing the Bottom-Up Approach as an Alternative to the Top-Down Approach for Dissemination

Researchers have used the top-down approach for dissemination. This approach applies efficacy evaluation to develop evidence-based interventions (EBIs) and disseminates them for community-based organizations (CBOs) to use. EBIs are rigorous but do not address feasibility and acceptance issues. CBOs are reluctant to use EBIs. TDE proposes the bottom-up approach as an alternative to overcome the limitation (Chen, 2010). The bottom-up approach starts with viability evaluation to ensure a proposed intervention is viable in a community. If an intervention's viability is identified, effectuality evaluation and generalization evaluation can be followed for providing evidence on the intervention's real-world performance for dissemination consideration.

CONCLUSION

This chapter has introduced TDE's concepts, approaches, and applications for serving stakeholders' evaluation needs. The introduction covered the conventional applications of TDE in assessing the implementation processes and outcomes. Furthermore, this chapter has discussed TDE's novel contributions in the following areas: (1) developing innovative proactive evaluation approaches; (2) providing a real-world approach for outcome evaluation

as opposed to the internal validity-focused outcome evaluation; (3) providing a valid approach for articulating situations, contexts, and circumstances of communities for a culturally responsive evaluation; (4) providing an innovative approach for assessing transferability as an alternative to the external validity approach; (5) expanding the evaluation scope by examining human factors in interorganizational collaboration; (6) examining relative strengths and limitations of using a behavioral theory versus stakeholder theory in an intervention program; and (7) proposing the bottom-up approach as an alternative to the top-down approach for dissemination.

REFERENCES

- Bledsoe, K. L., & Donaldson, S. I. (2015). Culturally responsive theory-driven evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 3–27). Information Age.
- Checkel, J. T. (2008). Process tracing. In A. Klotz & D. Prakash (Eds.), *Qualitative methods in international relations: A pluralist guide* (pp. 114–127). Palgrave Macmillan UK.
- Chen, H. T. (1990). *Theory-driven evaluations*. SAGE.
- Chen, H. T. (1997). Normative evaluation of an anti-drug abuse program. *Evaluation and Program Planning*, 20(2), 195–204.
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. SAGE.
- Chen, H. T. (2010). The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning*, 33(3), 205–214.
- Chen, H. (2015). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2nd ed.). SAGE.
- Chen, H. T., Morosanu, L., Bellury, L. M., & Kimble, L. (2019). Multiwave formative evaluation of a retention program for minority nursing students: Intended effects, unintended consequences, and remedial actions. *American Journal of Evaluation*, 41, 71–88.
- Chen, H. T., Morosanu, L., Powell-Threets, K.,

- Lian, B., & Turner, N. (2019). Assessment of the monitoring and evaluation system of a population-based program: Theory-driven evaluation approach. *Evaluation and Program Planning*, 77, Article 101719.
- Chen, H., Pan, H.-L. W., Morosanu, L., & Turner, N. (2018). Using logic models and the action model/change model schema in planning the learning community program: A comparative case study. *Canadian Journal of Program Evaluation*, 33(1), 49–68.
- Chen, H. T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces*, 59, 106–122.
- Chen, H. T., & Turner, N. C. (2012). Formal theory versus stakeholder theory new insights from a tobacco-focused prevention program evaluation. *American Journal of Evaluation*, 33(3), 395–413.
- Chen, H. T., Yip, F., Lavonas, E. J., Iqbal, S., Turner, N., Cobb, B., & Garbe, P. (2014). Using the exhibited generalization approach to evaluate a carbon monoxide alarm ordinance. *Evaluation and Program Planning*, 47, 35–44.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Scholer, D. (2011). A system review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–266.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Erlbaum.
- Fulbright-Anderson, K., Kubisch, A. C., & Connell, J. P. (1998). *New approaches to evaluating community initiatives: Vol. 2. Theory, measurement and analysis*. Aspen Institute.
- Hood, S., Hopson, R., & Kirkhart, K. (2015). Culturally responsive evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 281–318). Jossey-Bass.
- Pawson, R., & Tilley, N. (2007). *Realistic evaluation*. SAGE.
- Rogers, P. J., & Weiss, C. H. (2007). Theory-based evaluation: Reflections ten years on: Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 2007(114), 63–81.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.



PART III

VALUING

CHAPTER 7

Valuing Science

Theory as a Guide to Practice

Michael Scriven

It is worthwhile to consider the various stages in our metatheory of evaluation in order to trace the conceptualization of a transscientific theory of evaluation. There are several levels at which one can talk about “theories of evaluation.” These run from (1) simple lists of things that need to be “done” (i.e., covered in an adequate evaluation exercise) to (2) detailed accounts of how a particular sociological or political or psychological perspective applies to the particular operations of evaluations (e.g., evaluation as an exercise of power or evaluation as an exercise in male sexism, evaluation as a further example of some other sociopolitical bias); and (3) evaluation as a metadiscipline, meaning (for me) evaluation as a discipline that springs from self-referential linguistic activities. Without going into details, this view is of great importance because the main difference between *Homo sapiens* and other species with languages seems to me to lie in the fact that only humans reach what one can think of as *sophisticated* self-consciousness, in that they begin to develop a critical language for evaluating primary

levels of self-consciousness (i.e., recognizing certain noises or marks as signals to behave or think in a certain way—for example, a command, acquiescence, hostility, appreciation). In the primitive process of teaching a skill, even at the level of labeling (i.e., classification), it is hard to find any nonhumans engaging in critical discourse. When these discussions become useful is when we begin to talk about better and worse linguistic communication (e.g., when we begin to introduce the ideas of useful descriptions vs. unnecessarily vague or lengthy, or obscure descriptions). From this theoretical point of view, evaluation is important in explaining what about the human brain accounts for the human dominance among the species in problem solving. But a discussion at that level is not the issue here.

In the history of evaluation, as a discipline, the first theory of evaluation was what we will call the X-type theory, which was the theory that evaluation—as a discipline—was a nonstarter (i.e., not a discipline at all, but simply a linguistic category that included statements of preference, taste,

or prejudice, for which no scientific basis beyond description was appropriate). Evaluation theory started with the view in the late 19th and early 20th centuries that evaluation was nonscientific or at best prescientific. After the dominance of the positivist view faded somewhat we worked for a while trying to meet the description of evaluation as a new science (i.e., the Scientific View), but then it became clear that this was only part of the story. The rest of the story was that evaluation was more than scientific in two ways: First, it transcended science in that its concepts applied with equal force to many nonsciences (e.g., history, athletics, computer games, and ethics). Second, since the concepts that dominate science itself—the search for good or better theories, explanations, and classifications—are themselves evaluative terms, the search for the logic of evaluation is a fundamental logical search that goes deeper than the particular searches referred to by these terms. So, there is a double sense in which evaluation theory is transscientific, and this realization is the central theme of much of my work.

Talking about the utility of theories of evaluation for practice is all about the less abstract A-type theory or, to a limited extent, B-type theories. But I try to pull a rabbit out of the C-type hat.

The first true theory of evaluation was the A-type theory from Dan Stufflebeam, commonly called—from the beginning—the CIPP theory, standing for Context, Input, Process, and Product. This has most recently been refined in his posthumously published book *The CIPP Evaluation Model*. I then published another checklist model, whose current version is available online at michaelscriven.info. It is called the Key Evaluation Checklist (KEC), not intentionally to invite comparisons with the CIPP model, but Dan has published many criticisms of it from the CIPP point of view, which have led to some improvements in later editions of the KEC model. I hope that I will be able to further improve it in its online version,

and I look forward to critical and competitive input from readers.

It will be clear from the preceding remarks that the gap between theory and practice for A-type theories should be negligible, since the A-type *theory* is a list of required *practices*. The problem for B-type and C-type theorists is much more serious but may be very enlightening.

I have also produced something that is often referred to as a theory, the so-called “goal-free model of evaluation.” This is a simple B-type theory, based on the need to break loose from the inherited tendency, going back to various theories of program management that make reference to evaluation only in the sense in which it is devoted to the process of determining the extent to which the goals of a program have been met. It is still an approach that is widely advocated in the literature of management training and is heavily responsible for the failure to treat bad side effects with the same care as achieved goals, which leads, according to me—to the so-called “general positive bias” in most published program evaluations, because they don’t search for unintended effects, especially bad ones, with the same care as goal effects. The more evaluations that I see, the clearer is the “general positive bias,” and for this reason, I think the discipline of evaluation, at this point, must take on the load of counteracting the “general positive bias.”

Returning to the comments about the A-type theories, including mine, the response to the editors’ request to spell out the connection between theories and practice will cause any A-type theorist to reply that making such a connection is exactly what an A-type theory does.

Except for a teeny-weeny problem, which is the gap between the language used in talking about the items that should be covered in our checklists and the precise language required for a satisfactory response to each test point in the checklist. For example, in the KEC, one of the checkpoints is

“Context,” which refers to issues such as the state of knowledge of the evaluators. It can hardly refer to the complete state of knowledge of the evaluators since this includes a vast amount of knowledge that is completely irrelevant to the current evaluation. So what part is relevant? I provide examples to clarify the notion of relevance here but do not attempt to give a general formula that will serve as an inclusion–exclusion algorithm in an artificial intelligence (AI) reconstruction. This gap between the headings in the checklists and the things to which the headings refer in a particular case of evaluation calls for some understanding of language and conceptual linkages that has never been thoroughly stipulated. We look at a particular place where this point gets us into trouble in a moment. However, the underlying complaint remains valid, since such points continue to crop up as we refine our understanding of the mildly theoretical concepts whose names turn up in the headings within a checklist. There is no ultimate level of explication that will eliminate all such future uncertainties. In my case, a very important point is hidden in the study of “Context,” which the whole 35,000-word unpacking of the KEC that you will find at the web address referred to earlier was not enough to shunt a very important issue off into the sidelines of irrelevance. We look at this in a moment, but, with this inclusion, I would argue for the need for only one more expansion of the level of detail. This—provided by Jane Davidson—in her outstanding *Evaluation Methodology Basics*. Remember that A-type theories cease to qualify as acceptable theories when they become too long to keep in mind when talking about competitive theories.

Now, we come to the last—and possibly the most important—point in the expansion of what counts as relevant knowledge of the evaluator or planner laying out the checklist. This concerns the following issue. Exactly what knowledge of ourselves, and in particular our brain contents and capacities,

should be included as relevant and hence identified when writing an evaluation of any project or plan, and so forth, that includes reference to the level of knowledge of any of those involved (e.g., as subjects, authors, or managers of the program being evaluated)? When we look for visible learning from some educational program, a very common interest in evaluating educational programs, we naturally focus intensely on changes in knowledge and self-knowledge about this knowledge on the part of students in the program. While it is obvious that we should look carefully at changes in that knowledge state in the area of knowledge addressed by the program, it is always equally important to look at knowledge acquired “on the side.” For example, educational programs about drugs are often the main source of knowledge that makes these drugs attractive, while also conveying some knowledge about their deleterious effects.

Now evaluations, in general, are meant to be, and frequently serve as, guides to action (e.g., about whether to use a certain program for teaching some subjects at a particular grade level in a particular school). Education about drugs would usually be regarded as unsatisfactory for teaching some subjects at a particular grade level in a particular school (e.g., education about drugs would usually be regarded as unsatisfactory if the rate of drug usage went up rather than down as a result of the program being evaluated). In general, evaluations can often have effects on behavior because they alter not only abstract knowledge levels but also specific behavior levels.

An extremely interesting example of this, not previously remarked upon as commonly as the adverse behavioral impact of “drug education” programs, concerns programs aimed at *ending* drug addiction. These programs about drug addiction (e.g., to smoking, drinking, or illegal drug using) are quite often persuasive about the possibility of remedial programs as frequently, but not universally, curative of undesirable

drug addiction. When provided in the last couple of years of secondary school, they are often followed by indulgence in heavy use of drugs, partly “legitimated” by the belief that any acquired addiction can probably be remediated. While about 20% of young adults who try these substances will never become addicted to them, the remainder may or may not be so fortunate, and many of these will not prove amenable to remediation of their addiction(s) by the conventional routes—Alcoholics Anonymous (AA) and so forth.

Now suppose that you are a young adult who has tried one of these addictive substances briefly but not to excess, and is considering his or her future. Should you regard yourself as still able to back out of addiction via one of the readily available treatments? Taking into account any familial evidence about addiction tendencies in the family and curative track records, too, it’s clear that you certainly cannot have any confidence that your own possible addiction will prove treatable by the usually available means. Given what you know about the probable long-term effects of addiction, it’s clear that the risk–reward analysis makes continued indulgence at this point highly hazardous. You may be able to get a better sense of the probabilities if your choice of psychotherapist is very successful—but this is not a trivial condition. Now why is this situation so important?

It is important because it is a common situation, and the social consequences of the wrong choice are extremely costly to society, as well as to many of the individuals concerned, and it is very important indeed as an analogy for an even more important decision about one’s future life by every person from roughly young adulthood to late middle age. I illustrate this vital importance by looking at some cases of escalating importance.

The case of mask wearing in a COVID pandemic, prior to substantial vaccination in his community (below 60%). Here, the situ-

ation is that if not vaccinated, a non-mask-wearer will possibly be an asymptomatic but infected potential vector and, even if vaccinated, may, on the evidence available in March 2021, still be an asymptomatic virus bearer. The choice for this person is thus between slight inconvenience and demonstrated rejection of advice, and possibly legal prohibition and sanctions at the possible expense of infecting and being infected. Such a person may believe that they are showing the importance of some (incorrectly alleged to be an unjustified) constraint on their behavior. It is therefore completely indefensible, although understandable by those with a low rating of the seriousness of COVID illness, and not even forgivable. It should probably be treated as approximately as undesirable as driving at 60 miles per hour on heavily iced and well-traveled roads.

The general message that is required by these examples is that covering the Context Checkpoint involves taking some account of what you and possible interaction with others will, or may, or can, do and what the scientific evidence is about the chances of your success with each of your options, and the probable cost of each option.

Now we come to the crucial question, the answer to which depends on the ones we’ve been discussing so far. When we are considering our options, which are the basis for judging the wisdom of our choices, we must not only look at the theoretical possibilities but the psychological possibilities for us as an individual, which may be considerably less in number. Which brings us to the famous \$64,000 question, which inflation has converted into the \$64,000,000 question: Should one adopt the ethical point of view throughout one’s adult life—which is approximately the same as, should one adopt an altruistic rather than an egoistic point of view to govern one’s choice of actions and attitudes?

People are probably born with highly variable empathic and sympathetic motivation,

ranging from a great insight into the feelings of others to absolute lack of concern for those feelings (as such, although there may be some indirect concern because those feelings may be considered important by somebody or some group whose attitudes and actions have direct importance to oneself). This indirect weighting of the importance of others is the reason for a thin shell of apparent altruism that often appears to be protecting a substantial group of those with whom one interacts, but it is, of course, no true guide to altruism if the person is affected only by the indirect pressure to take account of others' feelings. Now we are ready to address the question of whether there is a reason for the egoist to become an altruist.

There is one obvious reason for the addict to become free of addiction—namely, that they will then be able to escape the iron grip of the addictive substance and hence achieve a reasonable balance of their other vital goals with which the addiction interferes. If willing to try hard at living the kind of life that the current best practices of talk therapy and constrained living recommend, the addict has a significant chance. Why should the addict try to make this major change in their life, which we will now call a “deep shift” change?

The answer that I favor for its combination of brevity and potency avoids the complexity of game theory but, if I am right, makes the same point that game theory would support.

It is a two-stage answer. The first stage consists in explaining why a world in which everyone is altruistic will be a better place to live than one in which everyone is egoistic. This is so because almost everyone in any world sometimes encounters situations where a smaller number of people, by their own exertion or sacrifice, produce a significantly greater benefit for a larger number by undertaking a risk or task that faces the larger number. For example, the hospital workers or firefighters can save the health or lives of many more people than they make

up if they are willing to take on a greater risk than the group of prospective patients. This will result in many more deaths per hundred members of the sacrificial group than would be the case if they did not undertake the risk of this sacrifice, and that is why we should be extremely grateful to them for their altruistic commitment. Now the reasoning about this situation—the pandemic COVID situation—works just as well with any size group, short of the population of the planet. If two people become infected with a potentially fatal disease and one person at considerable—and optional—risk can make their demise unlikely, then altruism on their part will mean that two lives or life years will be saved instead of one, in the worst case. If the disease is never fatal but deleterious, then we will still be an advantage for belonging in groups where the smaller number can at some risk to themselves prevent the outcome of that risk from affecting the larger group by their own willingness to undertake preventive care or maintenance.

It seems reasonable to conclude that if you have a choice of people to associate with, even if that number is very small, you will on many occasions in a normal life benefit if the group consists of altruistic people. The most threatening alternative to this inference is the possibility that you can present yourself to your associates (and behave accordingly) as altruistic, whenever there is any chance that any deception may be uncovered—but when the risk becomes mortal, or nearly so, you save your own life at the expense of the others since their opinion and possible sanctions will no longer be more important than your survival. However, against this must be placed the facts that the “two-faced mode” will deprive you of an extremely extensive period of challenges involving no lethal consequences, but which would be rewarding for you if an honest altruist and frustrating if you are two-faced (i.e., in fact, an egoist). In the course of a normal lifetime, this would mean frequent and almost unlimited opportunities for really helping your friends

and being appreciated for so doing versus a possible solitary or dual occasion where you would save your own life at the probable expense of at least one close friend and your reputation.

To sum up, although you might think that the certainty of a life of appreciation by others for your thoughtful behavior toward them is no compensation for the possible loss of your despicable life, you must bear in mind that you are simply betting something certain and long lasting against a remote possibility that is unbreakably linked to a lifetime of nervous deception and attendant displeasure.

Think back to the situation facing an addict whose life has descended to the gutter. The person now faces the alternatives of indefinitely continued disgusting circumstances interspersed with short periods of ecstatic exhilaration versus a relatively brief tough period with no highlights but a continuing slightly better life and a reasonable chance for a very big upgrade—a rosy future—lying ahead. After trying both and failing, the two alternatives remain the same, although the probabilities for the rosy future have deteriorated. It is still the case that the “rosy future” choice is plausibly better.

The choice of altruism will be a less obvious winner as the years go by and the failures of therapy recur, but if the failures are very close, the diminution of motivation will be slighter and the absence of continued closure in life high but not ecstatic will persist, so the net effect may go the other way. Reflecting on these two cases—the egoist and the addict—at least makes it plausible to argue for taking the conventionally respectable and truly radical “deepshift” alternative as at least deserving sustained experimentation, including multiple repetitions.

I now want to add an important consideration to the weighing of alternatives in the case for altruism. Aristotle remarked, in a debate with an early determinist who was claiming that all our behavior could be explained in terms of factors over which we

have no control, that this was clearly false because, he said, we always have control over those with whom we fraternize and that choice leads to changing our characters in ways that are therefore under our control. Although a contemporary determinist would not be impressed by this argument, since it merely shifts the locus of control back one “generation,” there is a breath of truth in what he says because the vent of choice can be highlighted by the pro-Aristotelian in his account and his some commonsense way that improves the argument. Converting the point to a recommendation about modern addiction therapy with its heavy emphasis on environmental control and talk therapy, one must surely take seriously the possibility and defensibility of moral education following the path of addiction therapy (a rider on this line of thought suggests a similar line of compulsory detention plus therapy for mask-deniers and social distance violators in pandemic situations). In the K–12 context, the element of forced confinement is more or less absent, but the problem of displacing the assumed but unconstitutional primacy of religion as the basis of ethics preempts the possibility of widespread adoption of a merely rational approach, which brings us to the final comment I want to make about the rational basis of ethics.

Getting the rational basis accepted in the schools and in public life is facilitated if we consider an issue that so far has not been discussed in this chapter. It is, to my mind, a kind of swing vote in favor of the rational approach. So far, we have argued for that approach based on an analysis of treatments that improve our ways of handling disasters and normal life as it occurs to us all. The moment has now come to face the fact that humans are agents and not simply patients: We must look at what they can do by active participation in the alteration of the probabilities for which we are partly responsible. The allocation of empathy may begin as genetic, but its future configuration in the human psyche is heavily dependent on

the influence of siblings, parents, teachers, and other caregivers. Similarly, as Aristotle stressed, there is a great deal of mutual influencing in the flux of friends and work colleagues. It is important to modify our discussion of whom we choose to associate with by looking carefully at how we *choose* to change or be changed by our friends, colleagues, listeners, viewers, and readers. To make reasons dominant in the discussion of ethical issues is very important because it sets an example to others who are influenced by performance, rather than preaching, and provides the reasoning for those who listen for it. If we recalculate the probabilities talked about so far, in the light of the effects we can produce after convincing ourselves that reasoning justifies ethics, then

the case for taking the ethical path is greatly improved, not least because we have a very great amount of time to produce our effects on those with whom, including our families, we spend most of our waking hours.

To sum up, the big takeaways I have to offer come from the practical A-type KEC theory and the C-type metatheory that suggests the evaluative approach behind KEC leads to the conclusion that ethics is simply a necessity for inclusion in the most traditional program and product evaluation.

Which ethics? That's the next question. After all, the Ten Commandments have been said to support several contradictory claims about the ethics of war, divorce, gay marriage, and female priests. Even checklists need disambiguation. *Hasta luego!*

CHAPTER 8

Responsive Evaluation

Robert E. Stake
Merel Visse

INTRODUCTION

Responsive evaluation was christened in the 1960s when, thanks to Lee Cronbach, then president of the American Educational Research Association, asked aloud if a standardized approach should be recommended for the professional evaluation of education programs. Some of us, wide-eyed, said there is already too much attention paid to Ralph Tyler's (1942) call for focus on goal attainment. Shouldn't we organize around *what is*, more than *what ought to be*? We didn't realize that might put us on the values limb of Marv Alkin's (2006) tree. Let's think a while about what it means to have values-oriented evaluation.

VALUINGS

"I need to get those credits out of the way as soon as possible," a graduate student said. "It just takes me too long to get this degree." Another stated, "They should not allow us on campus. Some of us are health profes-

sionals and are vectors of the virus, exposed to COVID all the time." And a friend: "I am going to visit my mother in the hospital, despite the risk. I simply need to see her." Such remarks indicate what is important to people. Their statements, little and large, show what carries weight. Implicit valuings, most of the time, tacitly present in stories on how people are doing. Their stories reveal their relationship with the world as one of care. Care for a degree, care for their economic situation, care for risk of infection, care for someone. Their sense of accomplishment and well-being depends on credits, degrees, their health, their balance, and their relationship with family. Valuing, caring about something or someone, lies deep in their everyday existence.

Professional existence and work are infused with evaluation. It is not a matter of choice. One cannot do one's work without sensing the quality of things (Eiseley, 1971, p. 202; Persig, 1974). Health care, education, and social service are infused with evaluation. Improvements in the quality of care, teaching, and learning require refinement

in the recognition of merit and shortcoming (Scriven, 1967). People expect to have things evaluated. People grade teaching and health care delivery quickly and over time. When it comes to getting evaluated themselves and the things they do, people are apprehensive. They do not want to be seen as simple. Parents, administrators, and other stakeholders want high marks. They often do not think of evaluation time as an opportunity to make things better. Theirs is summative more than formative evaluation. Not because they choose it, but because the system forces it on them.

All of us make rational evaluations, but we also undergo more sensory appreciations of the world around us. Feelings and expressions evoked by politics, a good book, or an exciting movie. Being moved by a work of art. We are evaluative beings, sentient beings (Sayer, 2011), beings who appreciate, reject, accept. We think about the quality of our lives. We continuously care about what is good and bad. We do this all the time, even when sitting idle. In our professional lives, too, we share stories, vignettes, allusions, on how things are going, informally and formally, and we judge what needs to improve. We are all evaluators, sometimes as amateurs, other times as professionals. Valuing is our everyday existence. Not an isolated valuing, but a valuing closely connected with our sentient nature, our experience, not separated from facts and oughts, but entangled with our realities and imaginations. Distinctions between facts and values can be made, of course, but temporarily as a thought experiment, just like other artificial distinctions we make, such as between mind and body. We find *personal experience* to be the root of valuing, the source of responsivity, the predecessor of criteria, and we amplify it in this chapter.

META-EVALUATION

Sometimes we value things too highly. Some of an evaluating person's acuity comes naturally with age, and with challenge and

commitment, but unfortunately, just as naturally, increasingly it comes with diminution of skill and distortion. We accept or reject some valuing for the wrong reasons. Both indicate dependencies: We lost a sharp eye and now need to rebalance our attachments. Valuing that come from over-attachment or overidentification demand renewed autonomy, perhaps by objectification, perhaps by detachment, perhaps by reflection, but seldom by just letting go. Sometimes just a recalibration, for values are not fixed, but fluid and transform within the context of the relations with others, such as with stakeholders. Things matter to us, to programs, and organizations because they constitute us and them, they form the essence of a group or of our being, and we should honor them for that. All we address here is that valuing can be good or bad themselves, and therefore we need to continue preserving and protecting the practice of meta-evaluation. Over time, we work out how to find the middle ground between the extremist value positions (Aristotle, 1220b, 21–27).

With good meta-evaluation (i.e., with good management of evaluation), the eye becomes sharper and the will becomes stronger. We can train ourselves to be better evaluators. But ever and ever, some judgments go awry. Formal evaluation is the conscious disciplining of judgment. We train ourselves better to see value, but we mystify the fact that values are multifold, plural, depending on contexts. Each conscious step is taken to manage, to refine, to validate judgment extends its formalization.

Formalization is disciplined thinking, regularly a way of making thinking impersonal and artificial. We become more exact in identifying the things we evaluate, the evaluands. Soon we have styles, protocols, models, and mechanisms to refine our formalizations, and we argue about which are the better. But everyone who thinks that *tossing dwarfs* is conceptually offensive and should remain unsaid may be wrong-minded (Leget & Borry, 2009). In many

decisions on quality and criteria for good practices, the normative ideas dictate the practice and emasculate the context.

When the program is found low in quality, even if in a single remark, some people feel hurt and are likely to think that the evaluator is using inappropriate criteria or standards. They often think that the evaluator has less than full comprehension of what is being evaluated. They often have a good reason for thinking so. The evaluator is often insensitive to those persons' standards, preferring those that are rationally made explicit. Evaluators often act as if measurement of an easy criterion will represent the quality of the whole. Evaluation puts people at risk.

Of course, the experts and authorities in general agree, some but not fully. They see different aspirations and implications. They see the evaluand differently. They see formal evaluation as having purposes in addition to informing the highest-standing stakeholder. Regularly, evaluators have their own purposes to serve in giving high and low marks, such as displaying their expertise, furthering self-employment, and trying to upgrade the organization. Many evaluators are highly ethical, but they are not indifferent to their own well-being (House, 1980; MacDonald, 1999). The formal evaluation scene is often a political scene (Cronbach, 1977).

The well-being of the organization, community, and state is, to some extent, also at risk. An evaluation finding may strengthen or weaken working relationships and endorsements. It may cause shudders in power relations, sometimes more than shudders. But not usually, because the administrators somehow constrain the design of evaluation to limit institutional risk. Questions of management of the evaluand often are not included in the study. Many administrators think that competence for evaluating the evaluand is different from competence for evaluating administration. And administrators are skilled at putting a promotional interpretation on the findings.

Valuing is a complex matter. One should expect some of the better efforts of evaluation will be ones seeking out complexity.

VALUES ARE NOT INTRINSIC

The value of something is created by humans. Horses crave water and moths crave light, but it is the value set by humans by which we judge the quality of quenching and beckoning. Health and safety are important to all creatures, but people concentrate their attention on the value placed by people. There may be a comfort for horses that horses know, but the comfort for horses that we know is a human conceptualization. For teaching and learning, timeliness and usefulness are not intrinsic qualities; schooling is valued because humans find it so. People do not value an object the same in different circumstances. Food is less valuable after dinner; safety is less valuable in the absence of threat. The voice of a tenor is more cherished when singing before the diva soprano than after. The obligation of an evaluator includes description and interpretation of the situation. The merit and worth of the evaluand depend on the writer's writing.

Appetite for evaluation often diminishes over time. By the time an evaluator gathers evidence of value, the commissioner of the evaluation and the chief stakeholders (the students, the teachers, the managers) often have found new fish to fry. Even the evaluator has different perceptions of his or her data as the study moves along. Quality is a market quotation, fluctuating over time and place. Hence . . . some people who lack summary knowledge but have extended experience are needed for interpretation of program values.

RESPONSIVITY

Evaluation responds to something, fixated on an evaluand, the thing being evaluated.

It aims to see the value, the goodness, the shortcomings, the worth, and utility of the evaluand. But first, evaluation needs to see, and then to say, what it is that is being evaluated—not just its name, its pedigree, its diagnosis, but its doing. The readers of reports will be misinformed if they are not given description of what it was the evaluators were responding to.

What, in 1973, we decided to call *responsive evaluation* is an orientation or approach that responds to the performance, function, situation, and accomplishment of the evaluand. It responds particularly to the experience of the people engaged—and, of course, to the merit inferred or somehow graded and declared.

All program evaluation is responsive to some mediation of criteria and standards, responsive to history and political surround. The characteristics of responsive evaluation described here are to be found to some extent in almost all evaluation contracts, regardless of method. All evaluation is shaped by the intents and outlook of the evaluators. Responsive evaluators try hard to draw their findings, their probative evidence, as Michael Scriven (1994) would have it, mostly from the activity and experiences within the evaluand.

Experience is emphasized. Criteria are largely *derived* from what is observed. Many evaluation methods identify goals, problems, or criteria, as a starting point for the study, as a design. Responsive evaluators do not ignore intentions and aspirations but prefer to let unstated criteria emerge from long and close acquaintance with the operation of the evaluand.

If they cannot themselves see the operation, evaluators depend on others, those who do, those in touch, those being touched. The running evidence of good happenings and hurtful traumas can be recorded in stories and artifacts, triangulated, and aggregated in judgments of quality by those with evaluation responsibility. It is not participant evaluation. It is not theory evaluation

or policy evaluation. Responsive evaluation is interpretive, dialogic, documentary, qualitative research, usually on a sociologically important human undertaking, the evaluand.

Responsive evaluation is particularistic, attending to the uniqueness of the mission, context, culture, and people of the evaluand. It may contribute to some general understanding of similar evaluands, but knowing this particular one is the task. Little effort is made to inquire or report how this evaluand stands against others. No situation can be well described without some attention to other situations, but the judgment of quality is made on the basis of goodness perceived, less on goodness compared.

The evaluation is subjective. Objectivity is never enough, even if that is all the contract calls for. Merit has to be interpreted, relying on responses of the staff, the beneficiaries, the administrators, and the sponsors. Those subjective data are to be examined critically; search is made for contradictory evidence. Measurements, questionnaires, and standardized interviews are used when helpful, and those results are also critically examined. The best brains of the evaluation team are expected to be active in teasing out the activity, implications, and values of the evaluand, recognizing future decisions to be made and commitments needing protection.

One criterion for the quality of evaluation is its utility, particularly as defined by Michael Patton (2008): How useful is the inquiry and report to people, not only to the sponsor or administrators but also to all stakeholders? Responsive evaluators, of course, are responsive to the contract and to the perceived needs of the stakeholders. How much stakeholders need the evaluation is a perception that changes during the evaluation period. But less changing is the value of the evaluand. The fundamental responsibility of the program evaluator is to find and report the goodness of the program.

HUMAN WORK

Responsive evaluation is said to be humanistic, meaning that it gives special attention to the work, the experience, and the valuings of people. In Education and Public Health, it is not just the views of syllabus writers, teachers, and students, but the values held by the evaluand's many stakeholders, and those further in the background, those of the stakeholders of other programs.

Work here is the strivings of the people of the evaluand, particularly those responsible for its conduct and service, particularly responsibility for its care and ethic. Not "work" as labor or as means to an end, but work as what we do that has an intrinsic meaning and that connects us with others—but also the strivings of the interaction of people to whom service is rendered, and to inhabitants of the closer and more distant communities affected by the evaluand. This work, the *vita active*, is communitarian work, and its geography needs to be scanned and held in sight.

Even for well-funded evaluations, there are few opportunities to survey community perceptions of need and desired accomplishment. It is up to the evaluator to be somewhat aware of the panorama of values, to know the better sources of relevant data, and to communicate with some who should know, such as writers and administrators. This overview cannot be as high quality a survey as that to be done of the activity and accomplishment of the program, but knowledge of background values, the values' context, needs to be realistic and pertinent.

We see responsive evaluation overview as an ethos more than a political topography. It is an ethos more than an ethic, where *ethos* refers more to the beliefs and practice of a culture of evaluation, and *ethic* refers more to moral principles. The responsive evaluator needs to recognize and to testify to the basic relevance of popular perceptions of the work being done and the work next needing to be done. Often, these perceptions are weak, poorly grasped. Sometimes

they are more visible in times of poverty and shortfall. When the program is struggling is a time for the evaluator to be sensitive—but not necessarily as a humanist. In times of stress, the evaluator should hew less to stated goals and doctrines.

This commitment to observed humanism does not include a general advocacy of humanistic philosophy. All professional evaluators should be advocates of the search for truth, the striving for fairness, and other moral pursuits, but taking a political or academic stand is not part of the definition of responsive evaluation. Its low reliance on prescribed goals, mandated standards, and uniform codes of ethics is based on their poor recognition of complex problems and defects in human relationships (Stake, 2004), especially where conflicts of principle are found.

The brunt of responsivity is a fixed focus on the human experience being made in the program over a given stretch of time, with less attention to the intent at the beginning and the performance at moment's end.

BLAME

As presently it happens, program evaluation deals much with rounded numbers. Means, deviations, and trends are used to summarize quality. Rounding often gets us much too simple a story. The actual steps trodden forward and back are made by individuals mainly, but groups have a complexity, too, needing reporting. A concentration on central tendencies ensures that not much will be known about the patterns and pitfalls of merit and worth, process and product.

Blame, too, as well as creativity, durability, and heroism, is reported in the gross. Leadership and abandonment are described in sentences, sometimes paragraphs, sometimes in a single tenor all through the report. We evaluators are seldom noted for our nuances. We do not tell the truth, the whole truth, and nothing but the truth. The truth is not simple.

And blame usually gets a special omission, avoidance. It hurts. It requires extra measures of evidence. It is extra work. Reporting blame for poor planning or performance is an obstacle to getting the next evaluation contract, unless the client had already decided that the evaluand should be shot down.

Is blame what stakeholders need to know? Only rarely is it what an internal evaluator needs to try to tease out. Only rarely can it be teased out given the budget provided. Perhaps it is better to omit matters of blame than to tell them wrong.

But merit cannot be understood without consideration of faithful conduct of responsibility. Nor without consideration of lapses of responsibility. Where lapses contribute to hurt or inequity or fraud, their telling is an obligation within evaluation. But it is also an obligation to seek out and describe their embeddedness in context and circumstance. Not an easy job.

Blame seldom yields itself to scales and significant differences. Blame is understood through stories, at certain times, and places and personalities, in relationships. We learn about the oblivious and the vindictive through experience, seldom our own; usually we hear the telling of experiences of others. Responsive evaluation is girded to learning about and the telling of experience, some of it about underperformance as part of the overall complex performance of participants.

STORIES

A responsive evaluation progress report or final report is likely to take the form of a story, or perhaps to include a number of stories. This format gives ample leeway for emphasis on the passage of time, the fixedness of place, the process of problem solving and its failures, and above all, the personalities of those involved. Many stories told are told as a revelation of problems developing; many evaluation stories show the problems

well in place. Most literary authors work toward resolution of the problem; most evaluations conclude with the problems still holding on, even when an effort to overcome nears the heroic. Some writings work to be entertaining or intimidating; most responsive writings work to show the deep intransigence of institutional and societal forces.

Many evaluation proposals promise to find a causal relationship between inputs and outputs. That is almost an unethical offering. Programs seldom have simple cause-and-effect relationships, with little and mostly unreplicable results.

An example: In the late 1970s, with encouragement from President Carter and others, several U.S. school districts participated in Cities in Schools (CIS), a federal program to bring urban social services into alliance with the local school operations, and consequently with better access for parents to those services (Stake, 1986). The design of the evaluation focused on student performance, on in-school and out-of-school criteria. The student data were contracted for but the schools were inconsistent in their delivery. From lack of evidence of student change, the evaluator found the alliance unsuccessful, with the schools at fault, but also rebuked the U.S. Department of Education (DOE) for funding this large-scale proposal with so little chance of success.

The DOE scheduled a meta-evaluation under the direction of Robert Stake (an author here) requiring it to be a stakeholder evaluation, with primary stakeholders: the Cities in School program, three participating urban districts, and participating social service agencies at each site. Also important stakeholders were the students and their families, the evaluators, and the federal government. Stake's proposal not only emphasized stakeholders but also identified this meta-evaluation as a responsive evaluation. On-site observation and interviewing provided much of the data. The work is described in Stake's 1986 book, *Quieting Reform: Social Science and Social Action in an*

Urban Youth Program. Illustrating the situationality, personalism, and contextuality is the following excerpt of an interview with one of the program's tutors, Sarah Davis. The combined data were interpreted to indicate the quality of the evaluation.

Interview with Sarah Davis, Student Computer Information System (CIS) Tutor

STAKE: We are talking about the CIS record-keeping system installed the second year. It wasn't part of your work the first year, right?

DAVIS: That's right. I went along with the CIS management information system enough to assure myself that I was accountable. Some people took it as a joke. Everyone complained about it. There were sheets for achievement, and sheets for social development, and others.

STAKE: Didn't you feel that information needed to be recorded?

DAVIS: Not all of it applied. My youngsters needed help with math. I didn't do much besides tutoring. The sheets may have been appropriate for teachers but weren't for us tutors.

STAKE: So you objected? Why?

DAVIS: Well, it was burdensome. It didn't seem important, partly because of the irregularity. Sometimes they had to have them the next day; sometimes not until the next week. Sometimes a group leader would treat them as important, other times not.

STAKE: So sometimes you kept them and sometimes you didn't?

DAVIS: No. I always kept the parts pertinent to the tutoring. Some of us did, others did not. A lot of them were not professional about anything. They didn't know what it was all about. Some didn't care.

STAKE: About the record keeping, or about the kids?

DAVIS: Some didn't care about either one. But the unprofessional ones mostly just

did things their own way. And some of it was pretty bad.

STAKE: What do you mean?

DAVIS: Well, they would manipulate kids. They probably didn't hurt anyone. They did cause a division in the staff. Some were pretty racist. I'm sure some weren't aware of what they were doing. Mostly they were not enabling the kids to grow—making the kids more dependent, not less dependent, but more dependent on them.

STAKE: Wasn't there training for them (teachers and tutors)?

DAVIS: Very little.

STAKE: Did they falsify records?

DAVIS: Yes, it wasn't uncommon. People cheated.

STAKE: Are you sure? Did you see it yourself?

DAVIS: Yes.

STAKE: I heard it said that some of the team needed the social services more than the kids. Is that possible?

DAVIS: I think it was true.

STAKE: Do you think any of them were actually out to sabotage the system?

DAVIS: Some talked about boycotting it. Don't get me wrong. There were some really good people working in the program. The person I worked with was exceptional. He would say, "I'm doing my job. I don't need to make a home visit that many times a month. They can fire me." Sometimes he did crazy things, but he provided the resources a child needed.

STAKE: Was he offended by the evaluation?

DAVIS: People I was with did not really come in contact with the outside evaluators.

STAKE: Did you know [name deleted] got lots of credit for getting the management information system going? Was that well done?

DAVIS: I heard about it. I thought it not very consistent . . . lacking reasons for some of the things done . . . working under pressure. But the biggest pressure seemed

to be to get something going that would make [name deleted] look good.

STAKE: So people who saw it that way were not highly inclined to make the record system work?

DAVIS: I had mixed emotions. A lot of people did not understand that they needed to be accountable in some written form.

Social science encourages a penetrating, dispassionate search for differences, for causes. Its evaluations are nurtured to look for blame, for causes (see Tolstoy, in *War and Peace*, asking why an apple falls). Here, too much had been promised.

Cities in Schools then slowly disappeared. Who failed? The designers of the record-keeping system? The dreamer who thought of aligning social services with the schools? President Carter? Sarah Davis? Turning away from criteria of success, turning away from blame, responsive evaluation, as meta-evaluation, looked at the evaluation and the evaluand, for their experiences. The report gave its readers a vicarious experience of the value of the program from September 1978 to May 1980, in schools in three urban school districts. There were feelings of blame, and feelings of accomplishment, meriting a plea for thoughtful review of values and adding to the experience of efforts to improve the plight of schools and other social agencies.

Sarah Davis had part of the answer. We never found out the value of President Carter's social services (CIS in three cities) because, as the evaluator reported, the pipeline lacked input.

BRANCHES OF RESPONSIVE EVALUATION

Since our first nascent branching, and even before, responsive approaches have appeared in a variety of arbors, some, of course, to other seedlings and trunks. Leaf work and branches shoot in a variety of directions. A roots metaphor is easily over-worked.

Any qualitative evaluation might be called "responsive" were it to lean toward ethnographic case-based design and away from scalar and stochastic mensuration (Guba & Lincoln, 1981; Stake, 1983). From the beginning, our responsive evaluation was a mixed-model approach to assessment (Greene & Caracelli, 1997). MacDonald's (1993) Democratic Evaluation was a contemporary and sometimes collaborating approach.

One variation, culturally responsive evaluation, is a vigorous version of responsivity if we look at projects and publications. This interpretation originated at Arizona State University and the University of Illinois, prioritizing the cultural context of program and policy, and especially working toward social justice (Hood, Frierson, & Hopson, 2022). Some writers connected responsive evaluation to theory, such as new materialism (Freeman, 2018) and care ethics (Stake & Visse, 2021; Visse & Abma, 2018; Visse, Abma, & Widdershoven, 2015).

In Europe since the 1990s, especially in Scandinavia and the Netherlands, scholars at several universities (e.g., the University of Amsterdam and the University of Wageningen) adapted the responsive approach to interactive remediation. Some connected responsive evaluation to democratic evaluation (Hanberger, 2018). Others infused the approach with methods to gather and advocate stakeholder's perspectives by participatory action-research and co-creation, aiming for empowerment (Abma et al., 2019; Baur, Elteren, Nierse, & Abma, 2010; Woelders & Abma, 2019). Some fused responsive evaluation and fourth-generation evaluation to monitor and foster social innovation (Guba & Lincoln, 1989).

With this small summary, we hope to show a few alliances and differences in rationale. Each has its special sensitivities, and all depend on contexts and the stakeholders involved. With the present far-reaching concern for the quality of teaching, and for formative purposes, some evaluation offices are moving away from the view of evaluation for normative judgment toward

formative evaluation—even though still favoring standards-based rubrics (Arbesú, et al., 2023; Stake, Contreras, & Arbesú, 2012). Formative and summative purposes continue to be different, with different uses, data needs, and appeals to the teachers to be evaluated. Seeing a need for administrative control, most administrators stick with the summative and remain scarcely receptive to the advocates of the formative view.

Across all variations of responsive evaluation, the central thought calls for beginning with *what is* the experience, for whom, when, and in what context, and at the end interpreting the many *expressions of worth* (Stake, 1975, p. 14; see also Abma & Stake, 2002). The way we try to do it is to begin with being present, wearing an ethos of care, particularly for those whose work and life are put at risk. We practice discovering and remaining sensitive to the small irritations and the deep issues in their circumstance and try to learn about their world by seeing and hearing them. Even small words tell us what is valued. We seek useful insights. We honor context, cases, and care. There is much to look forward to in responsive evaluation as part of efforts to humanize programs and policies. Especially during a long-term recovery.

Were he still with us, Cronbach might wish still that there was a greater standardization of methods in program evaluation. But from a values perspective—in a diverse and conflicted society—it seems that evaluation methods should best be tailored to fit the situation.

REFERENCES

- Abma, T., Banks, S., Cook, T., Dias, S., Madsen, W., Springett, J., & Wright, M. T. (2019). *Participatory research for health and social well-being*. Springer.
- Abma, T., & Stake, R. (2002). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation*, 92, 7–22.
- Alkin, M. (2006). *Evaluation roots: Tracing theorists' views and influences*. SAGE.
- Arbesú, M. I., Diaz Barriga, F., Garcia, J. M., Rivera, A., Stake, R. E., & Visse, M., (2023). *Evaluation of teaching in higher education: New challenges, new realities*. Universidad Autonoma Metropolitana.
- Aristotle. (reprinted in 1980). *The Nicomachean Ethics*, 1,141b, 15ff.
- Baur, V., Elteren, A., Nierse, C., & Abma, T. (2010). Dealing with distrust and power dynamics: Asymmetric relations among stakeholders in responsive evaluation. *Evaluation*, 16(3), 233–248.
- Cronbach, L. (1977, April). Remarks to the New Society. *Evaluation Research Society*, 1, 1.
- Eiseley, L. (1971). The mind as nature. In *The night country* (pp. 195–226). Scribner's.
- Freeman, M. (2018). Dialogue, difference and care in responsive enactments of a world becoming. In M. Visse & T. A. Abma (Eds.), *Evaluation for a caring society* (pp. 185–207). Information Age.
- Greene, J., & Caracelli, C. (1997). *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms: New Directions for Evaluation*, 74. Wiley.
- Guba, E., & Lincoln, Y. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. Jossey-Bass.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. SAGE.
- Hanburger, A. (2018). Democratic caring evaluation for refugee children in Sweden. In M. Visse & T. A. Abma (Eds.), *Evaluation for a caring society* (pp. 83–105). Information Age.
- Hood, S. L., Frierson, H. T., & Hopson, R. K. (Eds.). (2022). *Race and culturally responsive inquiry in education: Improving research, evaluation, and assessment*. Harvard Education Press.
- House, E. (1980). The logic of evaluative argument. In *Evaluating with validity* (pp. 67–96). SAGE.
- Leget, C., & Borry, P. (2009). Nobody tosses a dwarf: The Relation between the Empirical and the Normative Reexamined. *Bioethics*, 23(4), 226–235.
- MacDonald, B. (1993). A political classification of evaluation studies in education. In M. Hammersley (Ed.), *Social research: Philosophy, politics and practice* (pp. 105–108). SAGE.
- MacDonald, B. (1999, November 9). [Statement on occasion of his investiture as Doctor Honoris Causa], University of Valladolid, Spain.

- Patton, M. (2008). *Utilization-focused evaluation*. SAGE.
- Persig, R. (1974). *Zen and the art of motorcycle maintenance*. George Morrow.
- Sayer, A. (2011). *Why things matter to people*. Cambridge University Press.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Rand McNally.
- Scriven, M. (1994). The final synthesis. *Evaluation Practice*, 15(3), 367–382.
- Stake, R. (1975). *Evaluating the arts in education: A responsive approach*. Merrill.
- Stake, R. (1983). Program evaluation, particularly responsive evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models* (pp. 287–310). Kluwer-Nijhoff.
- Stake, R. (1986). *Quieting reform: Social science and social action in an urban youth program*. University of Illinois Press.
- Stake, R. (2004). *Standard-based and responsive evaluation*. SAGE.
- Stake, R., Contreras, G., & Arbesú, I. (2012). Assessing the quality of a university—Particularly its teaching. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education* (pp. 3–14). Routledge.
- Stake, R., & Visse, M. (2021). *A paradigm of care*. Information Age.
- Tolstoy, L. (1869). *War and Peace*. (N. H. Dole, Trans., 1899). The Russian Messenger.
- Tyler, R. (1942). General statement on evaluation. *Journal of Educational Research*, 35(7), 492–501.
- Visse, M., & Abma, T. (Eds.). (2018). *Evaluation for a caring society*. Information Age.
- Visse, M., Abma, T. A., & Widdershoven, G. A. M. (2015). Practicing political care ethics: Can responsive evaluation foster democratic care? *Ethics and Social Welfare*, 9(2), 164–182.
- Woelders, S., & Abma, T. A. (2019). Participatory action research to enhance the collective involvement of residents in elderly care: About power, dialogue and understanding. *Action Research*, 17(4), 528–548.

CHAPTER 9

Deliberative Democratic Evaluation

The Evaluator as Moral Fiduciary

Ernest R. House

INTRODUCTION

I have long advocated the idea of justice in evaluation and, in particular, the deliberative democratic approach. As society changes, we might add new techniques and practices to meet new challenges and strengthen our approach. What might these challenges be, and what should we do to meet them?

Evaluations don't seem as effective as they once were. Evaluator Tom Schwandt (2019) contends that evaluation has entered a post-normal phase similar to that of postnormal science. In normal science, scientific studies are widely accepted by the government and public, and often acted upon. In postnormal times, findings are frequently disputed, criticized, and ignored. Environmental science is an example, as is the epidemiology of COVID-19.

I believe these changes in evaluation and science reflect changes in the larger society. Over decades, American society has become increasingly fragmented, polarized, distrustful, unequal, and corrupt (House,

2020). In such a society, evaluations become less effective. How can we make them more effective? Evaluators might act as moral fiduciaries, cultivate cognitive empathy, focus on deep stories and deep values, control for biases, and be transparent. In this chapter, I outline significant changes in American society, delineate several practices that can strengthen evaluations, discuss conditions of their implementation, and briefly tie these practices to evaluation theory.

SOCIETAL CHANGES

Over the past 70 years, American society has changed dramatically. When I was young, say about 1960, there was a consensus among the majority that the government was good and that society was on the right track. This consensus emerged from World War II and Roosevelt's New Deal (Kennedy, 1999). There was a palpable sense of unity and national purpose. During the 1960s, the consensus and trust in government began breaking down. There was the Vietnam War

and the Civil Rights movement, followed by Nixon and Watergate. These events engendered discord and distrust.

By 1980, Reagan could run on the slogan, “The government is not the solution to the problem; the government is the problem.” He initiated an era of privatization, deregulation, and strong antigovernment sentiment. In the new century, there was the contested Bush–Gore election, the trauma of September 11, 2001, the Afghan and Iraq wars, and the Great Financial Crisis. Distrust deepened. American society fragmented.

Three powerful trends emerged: a deep decline in trust, a sharp rise in inequality, and an increase in corruption. People did not trust the government, they did not trust their institutions, and they did not trust each other. Decline in trust was precipitous during Vietnam and Watergate, when trust in government dropped from 77% in 1964 to 30% by 1980 (Rothstein, 2018). By 2014, trust in government was 20%. Trust in Congress was 6%. Not all countries are like this. Trust in government in the Nordic countries is 70%.

Another powerful trend has been steepening inequality. In 1981, Reagan lowered the tax rate on the wealthy, followed by more tax cuts by later administrations. America now has the most extreme inequality of income and wealth of any developed country (Piketty, 2014; Saez & Zucman, 2019; Stiglitz, 2018). The three richest Americans own more than the bottom 160 million. The richest 0.1% have as much wealth as the lower 90% (Hacker & Pierson, 2020; Krugman, 2017). CEOs of major corporations used to make 20 times the average worker wage. They now make 360 times as much (Bebchuk & Fried, 2004). Top hedge fund managers make 100 million dollars per month (Saez & Zucman, 2019). Worker wages have stagnated for decades. Social mobility has declined. Nobel economist Joseph Stiglitz (2018) said the economy is rigged in favor of the wealthy. Conservative media mogul Rupert Murdoch noted that America is developing permanent social classes.

A third trend has been worsening corruption. When Governor Blagojevich of Illinois tried to fill the Senate seat vacated by Obama, few expected him to auction the seat to the highest bidder for personal gain. Patronage had become corruption. American society was beset with cronyism, favoritism, and conflicts of interest. Special interests dominated Congress and legislatures to secure special favors (Lessig, 2018). Seventy-five percent of Americans believe corruption is widespread throughout the government (Rothstein, 2018). Political scientists call such behavior *institutional corruption*.

Distrust, inequality, and corruption interact in downward spirals. The more a government is distrusted, the less it can remedy corruption or inequality, and the more it becomes vulnerable to cronyism and conflicts of interest. Distrust allows corruption that generates more distrust and inequality. These trends are reflected in dystopian novels, apocalyptic films, and polarized politics (Hacker & Pierson, 2020).

ACT AS A MORAL FIDUCIARY

How should evaluators behave in such a society? Problems in the larger society are manifested in evaluation. Evaluators encounter fragmentation, distrust, inequality, and corruption, such as conflicts of interest, in their work. They are often mistrusted and work with stakeholders with quite diverse views. Some sources of inequality, such as tax policies, lie beyond evaluation, but evaluators can attend to inequalities, fragmentation, distrust, and corruption within the evaluation space. They can act as moral fiduciaries. They might assume some moral fiduciary responsibility for helping those lower in the unequal socioeconomic structure. They can strive to protect the interests of those less advantaged. Here’s an example.

In 2009, 11,000 forensic rape kits were discovered in a police warehouse in Detroit. The kits had not been processed. For the victims, undergoing a rape investigation is invasive.

The process takes hours and involves swabbing samples from every orifice of the body. Rape is a crime often committed by repeat offenders, and rape kits provide an opportunity to identify serial rapists. Yet the kits remained unexamined for years.

In an examination of police files, Rebecca Campbell and her colleagues (Campbell, Shaw, & Fehler-Cabral, 2015) discovered that the police investigating the crimes repeatedly dismissed victim claims of rape on the grounds that the women were prostitutes, sexually permissive, or did not want their parents to know they had sexual partners. In the judgment of the police, the women's claims were not worth pursuing as crimes. Such judgments reflected police racial framing about the presumed character of minority women. Police reports often referred to victims in highly pejorative terms.

Campbell and her colleagues (2015) worked through the cases with police until the rape kits were processed, an exhausting exercise, with police officials admitting that classifying women this way was mistaken. In a sense, the researchers appealed to empirical evidence and the police sense of fairness. This was a difficult study to conduct because it involved conflicts of deep stories and deep values. During the study, investigators kept in mind the welfare of those who had been abused. Their welfare was most at stake, not that of the police or those who funded the study. In such conditions of inequality, evaluators might give priority to the interests of those less able to defend their interests.

What is a fiduciary? In the financial community, the idea of a fiduciary means that the financial agent must act in the best interests of the client. As an experienced investor, I can attest that investment transactions are riven with conflicts of interest. For example, if an investor asks the advice of a financial advisor, the advisor is free to recommend mutual funds that enrich the advisor, not the client. Advisors often receive fees or a portion of the investment from mutual funds they recommend. Cli-

ents suffer because this is not the best investment for them. This is standard practice in finance.

Fortunately, there are advisors who swear to a higher fiduciary ethical standard; that is, they offer advice that is in the best interests of the client. If they fail, they can be sued. Reformers have tried to make the fiduciary ethic the legal standard for investment professionals, but professionals have fought vigorously against this, an indication of how prevalent not acting in the best interests of the client is. Nonetheless, some advisors do choose to act as financial fiduciaries.

Evaluators might act as moral fiduciaries, meaning they act in the best interests of those less able to defend their interests rather than in the interests of the sponsors or other stakeholders. Such an ethic doesn't preclude acting in other stakeholder interests as well, but it does mean that those less advantaged will be given priority. In most programs, it's not difficult to determine who is less advantaged. In Detroit, Campbell and her colleagues (2015) chose to work through case records with the police to process the kits. They sought not to reform the police view overall, but to change their view about specific cases. Evaluators can take some responsibility for the severe inequality in their society by doing something about it in actual evaluations.

CULTIVATE COGNITIVE EMPATHY

In conducting the Detroit study, the evaluators exercised some complex skills. One was to understand what the police were thinking. That insight was key to figuring out why the rape kits had not been processed. In his analysis, Schwandt (2019) recommends that evaluators strive to understand other people's perspectives. Evaluators cannot assume that their perspective is the only perspective. This is critical in a fragmented society.

Sociologist Mario Small (2019) calls such understanding "cognitive empathy." He contends that social researchers are quanti-

tatively literate but not as qualitatively literate. Qualitative literacy entails understanding other perspectives in depth. Cognitive empathy is not feeling empathy. It's not feeling the same as others. Nor is it sympathy, feeling sorrow or pity for them. Rather, it's the ability to understand people's predicament as they understand it. Their view will seem rational within their perspective. We can understand why they believe the way they do. That doesn't mean we agree with them. Certainly, Campbell and her colleagues (2015) did not agree with the police stereotyping of minority women.

Small cautions against overgeneralizing other views. He calls this avoiding "out-group homogeneity bias." When groups are far removed from our perspective, their views appear less diverse than they are. Not all police are alike. Another caution is to be sensitive to data used as supporting evidence. For example, when journalists and others report standardized test scores, they often explain the results based on no empirical evidence whatsoever. They may contend that scores are low because the education system is dysfunctional. But there's no supporting evidence about causes in the test scores results. Where do such explanations come from? From beliefs and stereotypes held by those making the claims.

FOCUS ON DEEP STORIES

Deep stories are the subjective prisms through which we view the world, including how we feel about it. They are stories that "feel right" emotionally. They are interpretations of events and situations that people act on. The idea of deep stories comes from sociologist Arlie Hochschild (2016), who studied Tea Party voters in Lake Charles, Louisiana. They voted for Trump overwhelmingly in 2016. In their view, life is a long march toward the American dream, which lies just over the hill. But the line they're in is stalled. Others are cutting in line ahead of them, minorities and immi-

grants, people who used to be behind them. That's not fair in their view. Government agencies are helping these people. It's the government's fault. Trump also blames the government, minorities, and immigrants. He derides the elites. This is the "deep story" of the Tea Party. It's also the deep story of Fox News, their source of news. Fairness is critical, construed a particular way.

Similarly, sociologist Katherine Cramer (2016) studied people in small towns and rural areas of Wisconsin. Their towns are struggling, though people work hard. They believe the government is taking money away from them and giving it to minorities and immigrants in the big cities. In their view, decision makers don't respect rural people. These resentful voters tipped the state to Trump in 2016. People act on their deep stories.

There are other deep stories, like that of the progressives in Berkeley. In their view, Americans built a magnificent public square, but marauders invaded the square, dismantled it, and stole pieces to build private mansions. Massive accumulations of wealth threaten democracy itself. That's not fair, either. They supported Bernie Sanders (Hochschild, 2016). These are a few of America's deep stories.

If I were evaluating an environmental education program in these communities, it would help considerably to understand their deep stories. Lake Charles has oil refineries and toxic industries. Even though people die of cancer at high rates, the residents see these industries as vital to their livelihoods and have ways of thinking about the problem, including religion. In rural Wisconsin, my evaluation would be different, and different yet again in Berkeley.

CONCENTRATE ON DEEP VALUES

Deep values are entwined in deep stories. The Louisiana story is built around deep values about race, gender, and social class, around assumptions of a hierarchy of race,

gender, and social class. It's assumed that some should be behind others in the natural order, a vision derived in part from Louisiana history, with its tradition of populist movements led by politicians like Huey Long (Hochschild, 2016). Within this perspective, social status is determined by how far a person is from the bottom. Being ahead of minorities and immigrants is a matter of entitlement.

Everyone has a hierarchy of values, with some values more central to the belief structure. These might be called *deep values* or *core values*. For most, they include family, fairness, and ingroup loyalty. For some, they also include racism and sexism, deep-seated beliefs about who belongs where and deserves what. Not all deep values are good or benign. Everyone has a deep story and deep values, and it's worthwhile for evaluators to reflect on theirs. I've discussed mine elsewhere, which began in childhood as a Roosevelt Democrat (House, 2015). Evaluators might locate their story within the population of deep stories. Evaluators grow up in a particular region, social class, and identify with those in their vocation. Having an idea of your position helps you understand others and the constellation of perspectives in the evaluation space. An inability or reluctance to understand other perspectives is a common failure of American foreign policy. It can damage evaluations.

I'm not suggesting that all views are relative or equally good. Some views are better than others because they are more moral and soundly based. An informed view includes a grasp of other views. A moral view takes into account the welfare of others. Even in fragmented societies, there may be room for an agreement based on shared deep values. For example, in conducting an evaluation of environmental education in Lake Charles, I would focus on shared deep values. I would be unlikely to change their worldview, but I might find agreement about the value of environmental education in specific areas, such as how pollution affects

their children long term. They care deeply about their children's future, and they have a sense of fairness. They want to be fair to their children. The deep value of fairness plays a central role for everyone.

CONTROL FOR BIASES TO ENHANCE FAIRNESS AND HONESTY

Being biased means being influenced by things evaluators should not be influenced by. Biases might be technical, like sampling error, or social, like racial and sexual framing, or situational, like conflict of interest, or psychological, like inappropriate anchoring. Whatever the biases, they can result in distorted findings. Stakeholders are justified in seeing biased studies as unfair. Taking care to mitigate bias is critical in conducting fair evaluations. Fortunately, there are analyses about how to protect against different biases (House, 2011; Kahneman, 2011; Scriven, 1976; Shadish, Cook, & Campbell, 2002).

FOCUS ON RACIAL AND SEXUAL FRAMING

Racial framing is a key mechanism through which racist beliefs are perpetuated. They play a huge role in American history (DuBois, 1986; Feagin, 2013). The white racial frame asserts that African Americans are violent, criminal, unintelligent, lazy, and oversexed (Feagin, 2013, p. 101). Whites are superior. Whites are immersed in racial framing in childhood and often act on it unconsciously (House, 2017). The effects are pernicious. Something similar happens with stereotyping females. The tacit nature of racial and sexual framing makes such biases extremely difficult to eradicate. Evaluators should look carefully for social biases in programs and evaluations, including program effects (House, 2017). They should check their own predispositions, those of colleagues, and expect them to check theirs.

BE TRANSPARENT

Trust is at a premium in fragmented societies. Transparency engenders trust; lack of transparency engenders mistrust. In climates of distrust, people imagine bad things are happening. Evaluators should be transparent about what they're doing. In a Denver bilingual education program beset with years of distrust between the school administration and Latino community, in my role as a federal court monitor, I made clear to each group what I was doing. I was open to discussion and recommendations from them. I shared the data we collected and solicited advice as to what to collect next to determine whether the program was being implemented properly. Over time, transparency helped establish trust in the evaluation and improved trust among groups. Transparency is no panacea, but it helps.

IMPLEMENTING THE PRACTICES

The purpose of these practices is to enhance the effectiveness of evaluations and address societal problems. The practices, singly or collectively, could strengthen many evaluations.

Acting as a moral fiduciary would be most useful in projects where there is a steep hierarchy. Inequalities in the larger society are manifested in evaluations. This can occur in large or small, formative or summative evaluations. An early evaluation task would be to carefully consider the hierarchy. Race, gender, and social class are key. The criterion for success is how well the evaluation protects and prioritizes the interests of those least advantaged.

For example, in the Women Affirming Motherhood program, two women developed a program to provide prenatal and postnatal care to minority mothers (Alkin & Christie, 2019). Later, a large New York City foundation offered funding if the program became one of several demonstrat-

ing interagency cooperation, a foundation goal. There evolved a steep socioeconomic hierarchy extending from the mothers and infants to the chair of the foundation. This hierarchy included the mothers and infants, those working directly with the mothers and infants, the program founders, the coordinating agencies, the foundation officers, and the foundation board. The hierarchy itself presented problems of understanding and communication.

In such a hierarchy, the interests of the mothers and infants might well be neglected. Those in each position have different perspectives, values, and interests, as well as age, race, gender, and class differences. Since the program was developing, I proposed a formative evaluation focused on helping the mothers and infants foremost, those who were least advantaged. A central evaluation focus was to determine whether they were getting the services they needed. Even with the best intentions, their welfare might be overlooked amid the micropolitics and machinations of a large endeavor. We have seen the value of patient advocates who help patients in large hospitals. I envisioned the program founders, mothers, and core workers as prime audiences for the evaluation since they had the most effect on the infants and the most urgent need for information. (For details about how this might play out, see House, 2019.)

Cultivating cognitive empathy can be complementary to any evaluation strategy, large or small, and is perhaps more useful in formative studies. There are several ways to accomplish this, depending on evaluation resources. The most likely path is for the evaluator to grasp other viewpoints through extensive discussions. Cognitive empathy should make a difference in how the evaluation is designed and received. If not, the evaluator isn't delving deeply enough. Attaining cultural empathy can be built into the evaluation design by including regular interactions with stakeholders as ongoing data collection in the study.

Focusing on deep stories and deep values provides a sense of direction about what to look for. What are these people about? What do they want? How do they see the world? There is a logic to their actions *from their perspective*. Some glimpse of this logic is invaluable in designing and carrying out an effective evaluation. Evaluators don't need polished studies, such as those constructed by sociologists such as Desmond (2016), Hochschild (2016), and Kramer (2016). Generally, the more interactions evaluators have with stakeholders, the better the possibility of understanding.

There are a few deep values that bind us as a community, including family, fairness, and ingroup loyalty, as well as sexual and racial attitudes. How these are woven into deep stories marks us and leads us to act in certain ways. Since people share deep values, there is usually room for some cooperation, even amid fragmentation.

Controlling for biases is obligatory for all evaluations. By biases, I mean mistakes that can invalidate the study. If an evaluator draws an inappropriate sample, employs the wrong statistic, or uses racial or sexual stereotypes, observers might rightfully claim that such a study is biased, unfair, and even invalid. Over decades, evaluators and others have assembled lists of actions and conditions that can cause people to arrive at incorrect findings in certain conditions. We call these biases. Knowledge of biases is core knowledge in evaluation. All biases have been demonstrated empirically to cause errors when drawing conclusions.

How *transparent* should evaluators be? There should be no secret side deals with various stakeholders. Even if innocuous, these destroy trust. Explaining the study may be enough, but if some stakeholders are more interested in particular aspects of the study, evaluators should accommodate them. There is no need to explain every detail. Rather, tell them what they need to know and what they believe they need to

know. Interact with them frequently. Keep them informed periodically. Face-to-face meetings are most useful. People can "read" how credible you are through nonverbal cues. Also, you can learn much important information you never suspected. Transparency establishes trust in times of distrust.

CONNECTION TO THEORY

The ethic of the moral fiduciary fits a long tradition of moral philosophy. Rawls's (1971) theory of justice as fairness changed the dominant utilitarian conception of justice. With utilitarianism, you could justify fighting inflation by inducing a recession that forced large numbers of workers out of jobs. They would suffer, but the larger society benefited—the greatest good for the greatest number—but without regard for how benefits and suffering were distributed within society. In elaborate arguments, Rawls said this was unfair. We should attend to those less advantaged as a moral duty.

Later critics of Rawls said those left out of the decision processes should also have some voice in making decisions that affect them. They could not always rely on decision makers to make the right decisions without being included in the discussions. Deliberative democratic advocates recommended inclusion, discussion, and deliberation of stakeholders in decision making and evaluation (Gutmann & Thompson, 1996; House & Howe, 1999).

However, democratic participation alone doesn't seem potent enough in a society so unequal. Evaluators should take stronger actions to protect the interests of those less advantaged. Acting as moral fiduciaries is one way to help. Instead of balancing the interests of different stakeholders equally or putting the interests of sponsors first, evaluators might give priority to those who need help most. The core value is fairness, fairness in both society and evaluation.

REFERENCES

- Alkin, M., & Christie, C. A. (Eds.). (2019). Theorists' models in action. *New Directions in Evaluation*, 163, 11–18.
- Bebchuk, L., & Fried, J. (2004). *Pay without performance*. Harvard University Press.
- Campbell, R., Shaw, J., & Fehler-Cabral, G. (2015). Shelving justice: The discovery of thousands of untested rape kits in Detroit. *City and Community*, 14(2), 151–166.
- Cramer, K. (2016). *The politics of resentment*. University of Chicago Press.
- Desmond, M. (2016). *Evicted*. Crown.
- DuBois, W. E. B. (1986). *Writings*. Library of America.
- Feagin, J. R. (2013). *The White racial frame* (2nd ed.). Routledge.
- Guttmann, A., & Thompson, D. (1996). *Democracy and disagreement*. Belknap.
- Hacker, J. S., & Pierson, P. (2020). *Let them eat tweets*. Liveright.
- Hochschild, A. R. (2016). *Strangers in their own land*. New Press.
- House, E. R. (2011). Conflict of interest and Campbellian validity. *New Directions for Evaluation*, 130, 69–80.
- House, E. R. (2015). *Evaluating: Values, biases, and practical wisdom*. Information Age.
- House, E. R. (2017). Evaluation and the framing of race. *American Journal of Evaluation*, 38(2), 167–189.
- House, E. R. (2019). Evaluation with a focus on justice. In M. Alkin & C. A. Christie. (Eds.), *Theorists' models in action* [Special issue]. *New Directions in Evaluation*, 163, 61–72.
- House, E. R. (2020). Evaluating in a fragmented society. *Journal of Multidisciplinary Evaluation*, 16(36), 26–36.
- House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. SAGE.
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus & Giroux.
- Kennedy, D. M. (1999). *Freedom from fear*. Oxford University Press.
- Kramer, K. (2016). *The politics of resentment*. University of Chicago.
- Krugman, P. (2017). The gilded age. In H. Boushey, J. B. DeLong, & M. Steinbaum (Eds.), *After Piketty: The agenda for economics and equality* (pp. 60–71). Harvard University Press.
- Lessig, L. (2018). *America, compromised*. University of Chicago Press.
- Piketty, T. (2014). *Capital in the twenty-first century*. Belknap.
- Rawls, J. (1971). *A theory of justice*. Belknap.
- Rothstein, B. (2018). How the trust trap perpetuates inequality. *Scientific American*, 319(5), 7.
- Saez, E., & Zucman, G. (2019). *The triumph of injustice*. Norton.
- Schwandt, T. A. (2019). Post-normal evaluation? *Evaluation*, 25(3), 317–329.
- Scriven, M. (1976). Evaluation bias and its control. In G. V. Glass (Ed.), *Evaluation studies review annual* (pp. 119–139). SAGE.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Small, M. L. (2019, May 22). *Rhetoric and social science in a polarized society*. 2019 Spencer Lecture, American Educational Research Association, Toronto, Canada.
- Stiglitz, J. E. (2018). A rigged economy. *Scientific American*, 319(5), 57–61.

CHAPTER 10

Values-Engaged Democratic Evaluation

Evaluation and Democracy

Jennifer C. Greene

PRELUDE

I write this chapter from my position as an American evaluator who has evaluated programs in the United States, both public and private, that are intended to improve life quality for targeted peoples in particular places. I write from a quiet and unhurried place, initially created by retiring from my university job in 2018, and since then, ferociously nailed down by COVID-19. And I write from a commitment to our collective craft of evaluation and its justification in today's troubled world. I believe that evaluation is an important service to society. In our work, we engage meaningfully and respectfully with social and educational programs that themselves are designed and positioned to improve life quality, longevity, and meaningfulness for one or more targeted groups.

THE EVALUATION ENTERPRISE

From relatively modest beginnings in the middle of the previous century, the field

of educational and social program evaluation has become a vast global enterprise, encompassing multiple domains of human endeavor in both public and private sectors and for both nonprofit and for-profit endeavors. During these decades, evaluation has also incorporated multiple new methodologies and technologies that can address more complex questions and effectively extend our reach, especially to remote peoples and underserved communities. Accompanying this rapid growth of evaluation as a professional practice has been a proliferation of distinct evaluation theories and approaches, well represented in multiple evaluation handbooks in recent decades (e.g., Donaldson, Christie, & Mark, 2014; Donaldson & Piciotto, 2016; Newcomer, Hatry, & Wholey, 2015; Ryan & Cousins, 2009; Shaw, Greene, & Mark, 2006).

For this volume, the editors have invited selected evaluation scholar-practitioners to present their particular theory of evaluation in the form of a *guide to practice*. This is a grand ambition, primarily because most evaluation theories have little empirical

support for their hypotheses or their practical claims. The actual practice of evaluation happens in many different contexts for widely differing audiences, and with highly diverse evaluands, ranging from national health care initiatives to local recreational interventions for underserved youth. Generalizable claims about evaluation practice across these multiple differences in evaluation theory and context have been mostly out of reach for evaluation scholars.¹

Even so, this book is a worthy endeavor, as it gathers together in one place the wisdom and experience of diverse evaluation scholar-practitioners, gained from multiple cycles of implementing their preferred evaluation theory, reflecting on theory-to-practice accomplishments and limitations, revising their theory according, and then venturing back out to the field once again. After all, being in the field is where the challenges are the most acute and the rewards the most substantial.

On the Nature of Evaluation Theory, Briefly

In my view, evaluation theory today is a normative statement by an author or promoter of a particular theory that stipulates what evaluation *should* be, specifically, what questions and whose interests should be advanced by the evaluation, what evaluation approaches and methodologies are defensible and appropriate for addressing the particular questions at hand, which stakeholder voices and perspectives should be included in the evaluation, what criteria should be used to judge the quality and effectiveness of the program being evaluated, and, collectively across these commonplaces of evaluation, *what values should be advanced by the evaluation*. Table 10.1 presents these standard elements of an evaluation plan and its implementation. Note that values are often not explicitly stated in an evaluation proposal or plan; yet, because evaluation is a normative practice, values are always present.

To illustrate this perspective on evaluation theory, I briefly sketch the ambitions

TABLE 10.1. Core Elements of Evaluation Practice, and Thereby Theory

- Key purpose and audience
- Key evaluation questions
- Methodological design, including accompanying data collection instruments and samples, analysis plans, quality criteria for data
- Evaluation implementation
- Evaluation reporting and use
- Values advanced by this approach to evaluation

of two approaches to evaluating, at a state level, the Feeding America program, which is run by the largest domestic hunger relief organization in the United States.

- Policymakers in State A have requested an evaluation of Feeding America in State A as part of their budget planning for the next 5-year cycle. These policymakers are most interested in how many eligible people received food from Feeding America and how much food the average user received, compared to food needs standards. A survey design is recommended, with sampling to cover all areas of the state. A well-regarded survey firm will be hired to conduct this evaluation.

- Following the lead of policymakers in State A, policymakers in State B have requested the same evaluation of Feeding America in State B. State B policymakers further want to know how important the Feeding America food donation was to recipient families *and* how well particular groups of people who face chronic food shortages (e.g., the rural poor) were represented in those served. A survey plus selected case studies of targeted communities is the planned design for this evaluation study, likely to be conducted by university experts in these domains.

The evaluation plans for both State A and State B focus on each state's needs for information on the quantity and quality of food shared with food-insecure residents

of the state. This evaluation thereby serves to inform the family welfare and budgeting responsibilities of state policy and planning agencies, thereby advancing values of civic duty and responsibility. State B adds questions that engage, and thereby validate and affirm, the importance of *also* understanding diverse program participants' standpoints on and experiences with the program. This inclusion of experiences of program participants advances values of voice, equity, and respect for all peoples.

In other words, evaluation is *not* a values-neutral enterprise. Rather, evaluations as designed and implemented always address selected information needs and thereby advance the values and interests of some stakeholders and not others.

Values-Engaged Democratic Evaluation

Specifically, where do values show up in evaluation theory and practice? The short answer is just about everywhere. Evaluation is the social practice of collecting information about the quality and effectiveness of a given social program. In any given evaluation, how are the concepts of "program quality and effectiveness" defined, and by whom? And what information about program quality and effectiveness is collected? Just objective measures of outcomes? Or also narratives of participants' program experiences in context? And from which stakeholders are these data on program quality and effectiveness gathered, and which stakeholders are not asked about their program experiences? Answers to these and additional questions shape each evaluator's stance on the values that should be advanced in evaluation and, concomitantly, the evaluator's selection of an evaluation approach that advances the values chosen.

My own preference for democratic evaluation, envisioned as a social practice (rather than a scientific one) that is enlaced with values, is likely rooted in my coming-of-age during the rebellious years of the 1960s in the United States. My approach to evaluation was then further developed, refined,

and nourished over subsequent decades within an exceptionally vibrant and thoughtful international community of evaluation scholars and practitioners.² This stance is rooted in democratic ideals and values, especially including equity, fairness, liberty, and voice. This stance builds on the important work of Ernie House, Barry MacDonald, Helen Simons, and others on the key principles and values of democratic evaluation. Notably, to honor the equity commitments of democratic evaluation, the interests and concerns of intended program participants are foregrounded as *the* central focus for the evaluation. Attending to democratic fairness, the views and stances about the evaluand held by (ideally, representatives of all) other stakeholder groups—program leaders and staff, community partners, state politicians, and more—are also included and given thoughtful attention and consideration.

VALUES-ENGAGED DEMOCRATIC EVALUATION IN ACTION

This presentation now turns to more specific descriptions of what values-engaged democratic evaluation looks like in practice—not only an impossible task given endless variations in context, but also a critical task that constitutes the backbone of this volume. The text follows the editors' outline of four key issues to address. Also in response to the editors' request, for each issue, I offer a sampling of the kinds of practices or activities that would be observed if the evaluators were implementing the version of values-engaged democratic evaluation that I am sharing in this chapter.

What Are the Primary Goals of Values-Engaged Democratic Evaluation?

Values-engaged democratic evaluation aspires to advance the embracement and enactment of democratic ideals and values in the programs and contexts being studied. These ideals and values can include

fairness, equity, justice, voice, as well as community and generosity, among others. Democratic evaluators strive to design and conduct their evaluations with explicit and mindful attention to the concrete ways in which context-relevant democratic ideals may be advanced—both during the planning and implementation phases of the evaluation, and after the evaluation is completed. For example, in democratic evaluation, representatives of *all* key stakeholder groups should be included in the process of developing priority evaluation questions, constructing an evaluation design, and determining indicators of both positive and disappointing results. Key stakeholders, in rough order of priority, include the program participants and, as relevant, their families and friends; frontline staff and allies; agency staff in charge of the program; as well as policymakers and program funders with responsibility for the program, among possible others. By designing and conducting an evaluation with the explicit inclusion of the program perspectives and experiences of multiple program stakeholders—especially the intended program participants and frontline program staff—this approach to evaluation models democratic ideals in action.

Values-engaged democratic evaluation is an excellent match to program interventions that themselves are intended to advance democratic ideals and values. This confluence of agendas can generate results that well represent the context *and* that are actionable, and thereby particularly useful and defensible. In contexts where the evaluand is based on other-than-democratic values (e.g., efficiency, charity), a thoughtful values-engaged democratic evaluation could still generate useful results and perhaps catalyze meaningful discussions regarding the purpose and direction of the program that was evaluated.

Observables

If an evaluator or an evaluation team were implementing an *ideal* values-engaged dem-

ocratic approach to evaluation, an outside observer of an evaluation team meeting would certainly notice that team members include many different kinds of people. But the observer might not be able to easily identify which evaluation team members were the professional evaluators, which the politicians, which the program participants, and which the community leaders. Equity, respect, and parity would be the meeting norms. More practically, an observer of an evaluation team conducting a values-engaged, democratic evaluation could observe and record the following kinds of activities: regular evaluation team meetings that included representatives of all key stakeholder groups; team meetings held at places readily accessible by all key stakeholders; data collection activities conducted at times and places convenient to program participants, and supported and facilitated by key evaluation team members; and key stakeholder names listed as contributing authors to reports on the evaluation design, implementation, and findings; among others.

For What Kind of Evaluation Is a Values-Engaged Democratic Approach Most Appropriate?

In theory, a values-engaged democratic approach to evaluation could be appropriate for multiple kinds of program and evaluation contexts. The primary agenda of this evaluation approach is to advance democratic ideals and principles—across the varied stages of a program's development, implementation, and evaluation, as well as across varied programs, policies, politics, and contexts. This advancement of democratic values is enacted through the active and consequential contributions to the evaluation made by multiple stakeholders—especially program participants and frontline staff. So, the primary criterion for adopting a values-engaged democratic approach to evaluation is broad stakeholder support for the central role of democratic values and commitments throughout the evaluation process.

That said, in evaluation contexts, purposeful attention to democratic values and principles may well be especially influential at the *early stages of a program's development and implementation*, more so than when a program is relatively mature. Broadly, an evaluation of a new or newly revised program is often undertaken as a “formative” evaluation of program activities and processes, in contrast to a “summative” evaluation of program outcomes. For example, a formative evaluation of a new program can be undertaken to assess the contextual meaningfulness and practical value of various program components, as well as the appropriate fit of this program to the contexts in which it is implemented. Advocates and staff of a new program may also be eager for, or at least more open to, evaluative attention to the values being advanced in the program.

Furthermore, a values-engaged democratic approach to evaluation may be more appropriately and effectively designed and implemented for *small or modest-size program initiatives*, implemented in local or regional contexts, than for programs designed for large numbers of targeted participants, living in many different kinds of cities and communities. Effective values-engaged democratic evaluation requires time on site, opportunities for direct interactions with program staff and participants, excellent interpersonal and listening skills, and genuine respect for the dedication and work of the program staff. Yet, compared with the significant challenges of developing and implementing a meaningful and effective social or educational program, evaluation is a cakewalk.

In this section, I have labored to address its key question: *For what kinds of evaluation is a values-engaged democratic approach most appropriate?* With additional reflection, I offer the following thoughts. The values-engaged democratic approach to evaluation that I am advancing in this chapter represents a commitment to particular values—those of democratic ideals. This commitment gets enacted primarily through the process and interpersonal dimensions of an evaluation,

and only secondarily through the methodological and data-focused dimensions of our craft. So, *for what kinds of evaluation is a values-engaged democratic approach most appropriate?* My response is an evaluation in which space and time are purposefully allocated to thoughtful and respectful conversations among a diversity of stakeholders, in service of greater understanding and appreciation one of the other, of shared program commitments, *and* of a more democratically successful program.

Observables

Similar to my response to the first question, evidence of a values-engaged democratic evaluation in process would include the following: evaluation team meetings attended by evaluators, program participants and staff, and community leaders, with active and respectful participation by many; evidence of group norms of respect, equity, and parity; and a prioritizing of needs of the program being evaluated and the people implementing this program over the needs and expectations of the evaluation and the evaluators.

For Programs of What Size Is a Values-Engaged Democratic Evaluation Most Appropriate?

As suggested earlier, a values-engaged democratic approach to evaluation can likely best manifest its unique strengths and contributions in evaluations of small to medium-size programs. At the center of this way of doing evaluation are the relationships established between evaluators and diverse stakeholders, and the interactions that are anchored in these relationships. Of particular importance are the genuine opportunities provided for diverse stakeholders—notably including people who are program participants, frontline staff, and perhaps selected community leaders—to contribute their voice, experiences, and ideas to the collective endeavor that is the evaluation. This is certainly possible for a program of any size, but it is likely to be more influential and

effective in evaluations of small, community-based initiatives.

Observables

Again, as suggested in my responses to the first two questions, a values-engaged democratic evaluation would ideally invoke norms of respect, voice, equity, and caring. While such norms are often evident in the interactions and conversations of a group of people working on a collective task, actual interviews and conversations with evaluation team members could probe more deeply into character and meaningfulness of their evaluation experiences.

For Programs Addressing What Social or Community Needs Is a Values-Engaged Democratic Evaluation Most Appropriate?

A values-engaged, democratic evaluation would fit best for evaluating a program of direct importance to the local community, that is, a program that matters significantly to key evaluation stakeholders and citizens alike. Meaningful and effective democratic engagement in any process, including an evaluation process, requires commitment, insight, and a sincere willingness to listen well and work respectfully with others. A democratic process is anchored in respect for all voices and standpoints, including those that differ from or are even contrary to one's own.

If there are divisions and differences among the key stakeholders participating in the evaluation, a fair and open process should be enacted to resolve these differences. An appropriate starting point for such a process would be to specify the criteria that the evaluand must fulfill to be selected for the upcoming evaluation.³ Future plans and time lines could also be developed for evaluations of other valued local programs.

Observables

In many contexts and communities, deliberations over which local community program

to evaluate could certainly invoke strong opinions and emotions. As stated earlier, values-engaged democratic evaluations are anchored in norms of respect, equity, voice, inclusion, and caring. Such norms could usefully be invoked to guide these deliberations to a fair and satisfactory decision.

REFLECTIONS

The evaluation of educational and social programs is today a fixture in most countries' government-sponsored initiatives and programs. After all, governments are responsible for the well-being of their citizens and for the wise spending of public funds. Private foundations and companies also sponsor programs to benefit selected sectors of our communities, as well as fund evaluations of these programs and initiatives. Many of these government and non-profit evaluation sponsors would likely select a traditional mainstream evaluation approach as the best match for their information needs about their program's quality and effectiveness, especially for programs implemented on a large scale. In a mainstream approach, program success would be defined by attainment of stated program goals, on the average, for participant groups, compared to similar people who were not program participants. And this is a long-standing defensible approach to assessing program quality and success.

As an alternative to—and *perhaps also a partner with*—mainstream evaluation options, in this chapter, I have presented and illustrated a *values-engaged democratic approach* to program evaluation, following the issues framework set forth by the editors of this volume on evaluation practice. The primary agenda of this approach is to advance democratic values of equity, fairness, and voice as an explicit and integral part of the fabric of the evaluation, including the evaluation plan, processes, and products. While all evaluation approaches advocate for a particular set of values, I believe that democratic values are the most

justifiable and worthy values for our work as evaluators. In the ideal, I also believe that evaluation can be an educative process for varied program stakeholders, as they work together to assess the quality and effectiveness of their program.

In summary, the evaluation of especially public programs can make valuable contributions to a democratic government, as it generates fair and better understanding of the important merits and shortcomings of these programs. And evaluation theory remains a critical component of meaningful and effective evaluation practice. In my own evaluation practice, evaluation theory has nobly and effectively served as a guide to evaluation practice, a conversational partner, and a companion for diverse and often worthy adventures.

NOTES

1. Arguably, the primary exception to this claim is the considerable empirical research on actions that advance and support meaningful use of evaluation findings, as well as that docu-

mented in Michael Patton's considerable body of work on utilization-focused evaluation.

2. These also were the decades of rapid global growth and development of the evaluation enterprise.
3. One such criterion should stipulate the kinds of expertise and experience—among both the evaluation team and the local volunteer stakeholders—that are needed for the evaluation at hand.

REFERENCES

- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2014). *Credible and actionable evidence: The foundations for rigorous and influential evaluations* (2nd ed.). SAGE.
- Donaldson, S. I., & Piciotto, R. (Eds.). (2016). *Evaluation for an equitable society*. Information Age.
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (Eds.). (2015). *Handbook of practical program evaluation*. Jossey-Bass.
- Ryan, K. R., & Cousins, J. B. (Eds.). (2009). *The Sage international handbook of educational evaluation*. SAGE.
- Shaw, I., Greene, J. C., & Mark, M. M. (Eds.). (2006). *The Sage handbook of evaluation*. SAGE.

CHAPTER 11

Culturally Responsive Evaluation

Critical and Liberatory Roots

Rodney Hopson
Vidhya Shanker

Although there are some common denominators across all human experiences and groups, there is reason to believe that an overemphasis on the search for universals has been, at the very least, premature, if not mistaken. Indeed, in some cases, the search for universals has inhibited rather than enhanced the encirclement of social science knowledge. Despite the long history of this concern with the relationships between experience, behavior, and systems, insufficient attention has been given to the impact of cultural, ethnic, or gender experiences on the development of behavior and the social systems by which behavior is expressed. This neglect is probably the result of androcentric, cultrocentric, and ethnocentric chauvinism in Euro-American and male dominated production of social science knowledge. We refer to this chauvinism as communicentric bias: The tendency to make one's own community the center of the universe and the conceptual frame that constrains all thought.

—GORDON, MILLER, AND ROLLOCK (1990, p. 15)

BEING SOCIALIZED IN AN EVALUATION BOX WITH WHITE SUPREMACY ROOTS

The initial quote above by Edmund W. Gordon, Fayneese Miller, and David Rollock captures the communicentric bias of evaluation knowledge production that manifests through the seen and unseen, acknowledged and unacknowledged histories of evaluation as a field, a legacy that these authors and others in the social sciences and educational research have revealed over the last 30 years (Scheurich & Young, 1997;

Stanfield, 1985, 1999; 2011; Stanfield & Dennis, 1993; Zuberi & Bonilla-Silva, 2008). The article by Gordon and colleagues (1990), originally published more than 30 years ago in a leading journal (*Educational Researcher*), sets forth a challenge to rethink the inherent nature of communicentric, cultural, and methodological hegemony; the search for universals; and the neglect of knowledge bases that refer to and legitimize Black, Latino, and Native American scholarship. It closes with an invitation to join them in pursuing knowledge to advance liberation.

As two minoritized evaluation scholars and practitioners who labor to transform persistent patterns of empirically observable asymmetries in the social sciences, particularly evaluation, we note that the specific characteristics of colonialist, imperialist, caste-based, dominating, superordinating ways of knowing and being threaten our very genealogical and historical existence. Nevertheless, we accept both the invitation and the challenge proffered in the article, as well as the opportunity presented by this current (third) edition of *Evaluation Roots*—to interrogate the canon of evaluation, honor the field’s liberatory roots and critical aspects that continue to be erased and unknown, and illustrate the practical knowledge of culturally responsive evaluation (CRE).

Much like Schwandt (2008), we recognize “the dualism reflected in distinguishing theory from practice and knowledge from action (as well as mind from body, fact from value, etc.) . . . [as] broadly characteristic of the whole of the Western theoretical and scientific tradition of knowledge” (p. 30). We add that this distinction between cognitive knowledge and embodied knowledge is not only characteristic of knowledge systems derived from the European Enlightenment but is moreover classed, abled, gendered, and racialized in its application, wherein some people’s work is considered practice (e.g., service, caring, reproductive, farm, manufacturing, “unskilled” and often un/underpaid labor) while other people’s work is considered theory (e.g., management, leadership, governance, ownership, professional/ized labor). The knowledge required for and embodied within the former is dismissed and denigrated unless it can be articulated, explained, and written in a language literally and figuratively intelligible to the latter (Carden & Alkin, 2012). Indeed, this distinction has served as the gate that continues to keep out of books and the canon at large unconventional and innovative thinking about and practice of evaluation by members of the field who are excluded

from—and disproportionately lack access to—opportunities to conduct research and publish. The distinction also underlies the lack of commitment and incentive among some to conduct research and publish for audiences who do not necessarily share their sociopolitical experiences and interests in places and ways that are financially and linguistically inaccessible to the people to whom they hold themselves mutually accountable. As such, rather than presenting a series of performative checklists or mechanistic step-by-step techniques, our intent for this chapter is to explicate CRE as a distinctive *praxis*. CRE has its own structure and epistemology (Schwandt, 2008, p. 30)—a theory of practice and practice of theory—namely, a critical/liberatory tradition, linked through the iterative, dialogic nature of “response” and “responsiveness.” Taking Schwandt’s notion that “evaluation has its own unique kind of practical knowledge, comprising the tact, dispositions, and considered character of decision making called for in various situations faced in ‘doing’ the practice” (p. 33), this chapter thus claims and illustrates the practical knowledge traditions underlying CRE.

Much like Stanfield’s (1999, p. 426) declaration that “we all have been socialized in a social science box with White supremacy roots,” we acknowledge that our own training, orientations, and initial practices were in an evaluation field with biased theoretical work, narrow designs, and Eurocentric strategies that overwhelmingly continue to neglect racially otherized groups. Much of our own un/learning about evaluation’s roots took place through pursuit of each of our PhDs at Persistently White Institutions in the United States whose faculty have contributed to evaluation theory and practice. This un/learning has involved questioning the hegemonic theoretical and practical knowledges manifest in books like this and countering the “racialist and paternalistic traditions of social scientific work” that reproduces relations of “dominance and subordination . . . in the academy and

in the worlds [that] we study and evaluate” (p. 424).

I (Hopson) trained in the Curry School at the University of Virginia in the old Bureau of Educational Research in the mid-1990s, a stalwart of evaluation and educational research in the state of Virginia and among other emerging programs in evaluation at that time of its growth and development. Luminaries in the field were white-facing, men, and limited in evaluation models that incorporated recognition of cultural and contextual realities. I relied on coursework in linguistics and anthropology to supplement formal educational evaluation coursework and professional experiences related to cultural and linguistic attributes, ideologies, and frameworks. My major advisor, Robert Covert was a catalyst for thinking about the (multi-) cultural in evaluation professional settings and for introducing me to a small number of African American senior and junior scholars and practitioners—Anna Madison, James Earl Davis, and Sarita Davis, who pushed the boundaries of the field in the early 1990s (Madison, 1992) and/or were matriculating through other U.S. evaluation programs.

I (Shanker) came to the field in the mid-2000s from two decades of anti-oppression activism, organizing, and paid work. After having been laid off from a nonprofit addressing domestic violence that had lost some funding, I was simply seeking a credential that would draw on my talent with data from the nearby University of Minnesota. A certificate in evaluation would make both me and any future organization that I worked with more financially secure, I thought. My first course was with Jean King, who asked me to join her doctoral program. I was reluctant at first and started incrementally. In my assigned texts and at the professional association conference—even in sessions sponsored by the so-called “diversity” topical interest groups (TIGs)—I struggled to find any resonance and often instead found explicit “racialism and paternalism” (Stanfield, 1999). Through my own

search, I found a handful of texts that did name systems of oppression and challenge the naturalness of evaluation’s empirically observable racial hierarchy—by Bob Covert, Asa Hilliard, Anna Madison, James E. Davis, Kien Lee, John Stanfield, Hazel Symonette, Denise Ward Hood, Sandy Kerr, Veronica Thomas, Hayley Marama Cavino. I noticed that this work was rarely, if ever, cited or included in my syllabi, especially relative to that by authors—most recognizably white Women and sometimes Black men—who relied on vague language directed at the practitioner and methodological level.

Having internalized the disbelief that I had encountered when sharing my experience, I privately began to systematically track and code in an Excel workbook all the evaluation texts that signified racialized difference to test for myself whether my visceral sense of the literature’s trajectory—wherein critical analyses consistently stagnated, while liberal discourse proliferated—was “all in my head.” Was this a topic worthy of dissertation research? Or did I just have an axe to grind? I relied on the training I had gained outside evaluation to identify critical literature to develop my codebook. When I finally showed Dr. King 3 years’ worth of analysis of more than 5,000 texts, she said I had a dissertation. She courageously supported me whether or not she understood or was comfortable with my budding critique of the field’s invocation of diversity, cultural competence, and inclusion at the expense of a structural analysis of oppression and accompanying critical theory. This is in sharp contrast with other academics and employers who have gone out of their way to silence and punish what they could not predict or control by preemptively coaching attendees, sponsors, and discussants of conference sessions and other speaking engagements that I was part of. As recently as 2021 and as early as 2012, they discouraged discussants and others—explicitly and in writing—from entertaining my discussion of abolitionism, analysis of whiteness, or critique of cultural competence.

Without the perspectives of the authors who had advanced a structural analysis and critical theory of oppression in, through, and around evaluation, the research I had completed felt incomplete. I interviewed 12 pivotal authors to gain understanding of the factors that influenced the trajectory of evaluation literature signifying racialized difference (Shanker, 2019). While absent from all representations of evaluation's canon—both previous editions of *Evaluation Roots*, the Oral History Project, and the American Evaluation Association's (AEA's) theory awards—these critical theorists, disproportionately women of color and Indigenous women, had also been working with others whose writing was never published to recruit and retain evaluators from underestimated and thus underrepresented groups. They did so while building the field's understanding of oppression since AEA's inception without funding or fanfare. Grounded in liberatory movements, they have seemingly effortlessly connected the intrapersonal and interpersonal with the institutional and structural—"selves and systems," as Geri Peak is known to say—for decades, anticipating the systems turn in evaluation without having their theoretical contributions to evaluation recognized as such.

Awareness of these patterns of erasure and co-optation within the field led me and a group of women and non-binary-identified evaluators—largely Black, of color, and Indigenous—to collaboratively call evaluation's professional community to action.¹ Petition signatories demanded that the third edition of *Evaluation Roots* remedy its systematic omission of women of color and Indigenous women resulting from a narrow understanding of theory rooted in systems of domination and think more expansively about authorship and publication (Lopez et al., 2021). Exposure of the book's pattern of exclusion is only one piece of a much larger puzzle that echoes and amplifies the efforts of my foremothers within the field to build a critical/liberatory tradition in evaluation.

DEFINING AND SITUATING CRE AT THE THIRD EDITION OF *EVALUATION ROOTS*

CRE, at least in the current moment of the third edition of *Evaluation Roots*, seems firmly entrenched in larger systems of evaluation thinking and practice. CRE has earlier been defined as (1) an approach or model in a Stufflebeamian (Hood, 2001) sense used by evaluators and researchers to guide the quality and principles of evaluation; (2) "a system or culmination of practical strategies and frameworks that attend to culture during the various stages and phases of an evaluation, from preparing for it to disseminating and using the results" (Frierson, Hood, Huges, & Thomas, 2010); and (3) a theoretical, conceptual, and inherently political position that includes the centrality of and attunedness to culture in the theory and practice of evaluation (Hopson, 2009, pp. 430–431). Our purpose in this contributed chapter is to build on CRE's conceptual and theoretical underpinnings, extending its practical and praxis orientations in ways that bring it into the current evaluation roots discourse without defanging it of its inherent oppositional consciousness. While often unrecognized or not realized by those who obfuscate its origins or use its practice simply as framing, CRE's historical roots lie in the continual interrogation of and resistance against the socializing and hegemonizing techniques of evaluation within scholarly traditions and societal structures derived from European colonization and imperial expansion, as well as within racialized institutions and organizations (Ray, 2019). Indeed, as much as CRE requires responsiveness—a kind of call and response—with the cultures and communities potentially most adversely affected, it requires a response to white supremacy culture (Okun, 2001).

The Nobody Knows My Name project (Hood, 2001; Hood & Hopson, 2008, 2018; Hood, Hopson, & Frierson, 2005, 2015; Hopson & Hood, 2005)—a project that highlights early contributors and ground-

breaking African American scholars and practitioners excluded from the history of educational evaluation in the United States—documents the praxis of culturally responsive evaluation since before *Brown v. Board of Education* through historical manuscripts. Despite this history, entrée of the phrase “culturally responsive evaluation” into the academic halls of evaluation, and in fact into this book, is recent and deserves somewhat of a creation story. At a symposium in May nearly 25 years ago, a significant event occurred not far from where I (Hopson) currently teach in the U.S. Midwest. Two hundred fifty participants, speakers, and dignitaries from all corners of the earth were celebrating the illustrative career of Robert Stake, Professor Emeritus at the University of Illinois. While Stake needs no introduction, among the speakers—Cronbach, MacDonald, Lundgren, Bresler, Griggs, Denny, Kushner, Greene, House, Jaegar, Scriven, Smith (some of whom are on the second edition of the evaluation theory tree and all of whom are also familiar “household” names within evaluation)—was one Stafford Hood, former student of several of the distinguished Illinois faculty present and then-professor at Arizona State University. On this day, which took place not long after release of the film *Amistad*, Hood (1998) wondered aloud through his paper in Stake’s honor—titled “Culturally Responsive Evaluation as Responsive Evaluation *Amistad* Style”—whether “the *Amistad* incident . . . could serve as a lens [*sic*] . . . to better understand responsive evaluation and assist [him] and others in conducting culturally responsive evaluations” (p. 102).

Taking a page from the *Amistad* case before the U.S. Supreme Court in which Africans kidnapped from Sierra Leone were tried for mutinying all but two of their captors aboard the slave ship, Hood (1998) makes a couple of observations that situate the historical founding of CRE at such an auspicious moment. Lifting a few major points in the Stake Symposium that would reverberate with the practice traditions

underlying CRE, he summarizes the alignment between CRE and the *Amistad* case as follows: “One of the obvious similarities between the *Amistad* case and the evaluation of education programs is the participation of African Americans as experts in a professional manner that could decide the fate of the stakeholders of color” (p. 102). The practice tradition that this quote lifts is that of *participation*: Who is considered an expert, and how do we value experts in the evaluation endeavor? CRE’s respect, honoring, and privileging vis-à-vis subjugated knowledge and knowledge traditions is foundational. These notions of participation recognize the settler colonial mind-set that reframes the local, the Indigenous, the marginalized as caretaker of a (disproportionately white) communicentric world of data scientists, academic evaluators, and agents in the philanthropic and/or government sectors. “Participation” for these regimes tends to focus on involving communities in ways that fit outsiders’ interests, values, tools, methods, and ways of speaking about evaluation—usually for and under time considerations that require surface-level participation and involvement in meetings, convenings, and visits. The participation tradition of CRE is intended not to be episodic but continuous and longitudinal, not extractive but collaborative, not surface but deep, reflecting the ongoing relational aspects of the evaluand and attending to intersecting complexities of culture that exist in the diverse communities, contexts, and ecologies.

The parallels with respect to participation between CRE and the *Amistad* case extend beyond this preliminary understanding of participation, however. The captured Sierra Leoneans, destined to labor under conditions of enslavement and exploitation—wherein their language and culture would be systematically denigrated and denied expression—were not free upon the successful mutiny of their captors. Rather, they were to be punished for their acts of courage and were subjected to a trial requiring that they

defend their *deservingness*—another word for merit—of freedom from bondage. Avoiding punishment, and subsequent enslavement, put them at the mercy of U.S. jurisprudence as opposed to their own system of restorative justice, which not only would have been familiar to them but also would have reflected their indigenous cultural values, relational code of ethics, and spirituality. As a result, the captured Sierra Leoneans had to make their case legible in English as opposed to their tribal language of Mende or even the lingua franca of Krio to a jury that included none of their peers. Similarly, while CRE prioritizes participation, it operates within a larger system wherein labor is racially and socially stratified. At the mercy of commissioners of evaluation, editors, and “peer” reviewers, CRE practitioners must continually make their case. Typically, “the business case” (Weisinger, Borges-Méndez, & Milofsky, 2015) or the “diversity case.” It includes “multicultural validity” (Kirkhart, 1995) and involves using a foreign tongue, logic, and values that are not necessarily shared by those potentially adversely affected but that must be intelligible to those with decision-making authority: the ultimate judge and jury. Too often, the primary intended users of evaluation results—CRE included—remain outside the communities most adversely affected by the situation, program, or evaluation. Last, the captured Sierra Leoneans’ active role in ending their bondage did not result in the end of bondage for all. It did not even end—and could never have ended—the systematic capture, trafficking, exploitation, and enslavement of Africans and others. But the movie reinforced viewers’ trust and belief in U.S. jurisprudence as a neutral system and noble arbiter of justice. In parallel, the relative comfort that participants in and practitioners of CRE experience does not prevent anyone else from experiencing epistemic violence and epistemicide through evaluation—let alone repair, reverse, or regenerate from that harm (Shanker & Maikuri, 2021). CRE does not and cannot,

by itself, alter the hierarchical structure of evaluative knowledge production. Its existence on the evaluation theory tree can lull those in and around evaluation with a sense of satisfaction in the capacity of the field’s existing structure to produce the illusion of “social justice.”

A second quote from Hood (1998) explicates another important practice knowledge tradition in CRE—the recognition and acknowledgment of the benefits and focus of *lived experiences* as important cultural context markers. Hood’s quote is as follows and highlights the particular role of the interpreter and the processes by which one identifies and searches for interpreters:

A culturally responsive evaluation could include evaluators, observers, or interviewers who do not share the racial background of the culturally diverse group of stakeholder/participants. However, the extent of their lived experience in the cultural context of the participants and understanding of the groups’ verbal and nonverbal communication must be closely scrutinized. (p. 109)

The CRE tradition of lived experiences is by its nature both deliberate and necessary. Prioritization of lived experiences into the evaluation is a precondition of CRE and requires understanding the nature of lived experiences in ways that intersectionally accentuate cultural contexts and situations that involve evaluated-program participants, especially. This practice tradition of CRE involves ensuring that the lived experiences of those potentially most adversely affected are not engaged symbolically but rather are integral to the development of measures, tools, resources, frameworks, and approaches.

Again, the similarities between CRE and the *Amistad* case are multilayered in that uses and understandings of “lived experiences” typically racialize, and essentialize, culture without paying sufficient attention to intersectionality (Crenshaw, 1989) and the matrix of domination—the interlocking systems of racialized and gendered capital-

ism (Collins, 2000)—that shapes everyone’s lived experiences. Black Feminist Thought theorizes power beyond the interpersonal level, accounting for hegemonic, disciplinary, and structural power, all of which impinge on individuals’ experiences. These tend to vary more within artificially constructed classifications than between them. As a result, while interpreters and translators may share some dimensions of experience with those most adversely affected, they likely do not share all (Madison, 1992; Ward Hood & Cassaro, 2002). And as much as participation cannot be mistaken for structural change, interpretation/translation cannot be mistaken for participation. This is where the notion of responsiveness is pivotal. “Response” offers evaluators the opportunity to circle back with those most adversely affected on every decision—precisely because evaluators (and interpreters/translators) cannot assume that their lived experience or “culture” can suffice as a stand-in for participation among and collaboration with those most adversely affected, let alone that all those most adversely affected share the same experience. CRE practitioners bear a responsibility to hold themselves laterally accountable to those on both sides of the interpretation/translation and ensure that they facilitate intelligibility across power differentials (Peak, Peters, & Fishman, 2007; Symonette, 2004, 2015). Done symbolically, CRE reifies racialized, essentialized notions of culture, contributing to the ongoing commodification and exploitation of phenotypic, cultural, and linguistic difference within evaluation. But done self-reflexively (Fine, 1998) and with mutual accountability, CRE challenges the individualism on which not only positivist and scientific epistemologies but also carceral and meritocratic logics—including with respect to theory development and knowledge production—are based. It thus constitutes a challenge to white supremacist, cisheteropatriarchal, ableist capitalism, neoliberalism, and other systems of domination in the process.

BEYOND A CRE DESTINATION AND TOWARD A LIBERATORY ONE

While this chapter explains the creation story of “CRE” as a phrase, the roots of CRE as a praxis go back much farther than the phrase’s coinage in the late 1990s—at least to legislation resulting from the Civil Rights Movement, and perhaps even farther back to the onset of residential, occupational, and educational segregation and violence that necessitated the struggle for true, material emancipation after the Emancipation Proclamation. It is during this pre-Civil Rights era that African Americans are first documented by Hood (2001) as being credentialed in, publishing in, and practicing evaluation with and for their communities—after all, they were denied access elsewhere. Much the way that Cavino (2013) conceptualizes evaluation as necessarily furthering colonization or decolonization, depending on how it is approached, because of its association with the settler colonial state apparatus under conditions of occupation, educational evaluation in the context of legally (and “postlegally”) segregated education necessarily furthers emancipation or exploitation, depending on how it is approached. Hood (2017) documents this critical consciousness among the likes of Ambrose Caliver. Cavino and Hood thus describe precursors to what might now be called principles focused evaluation (Patton, 2017)—evaluation that serves the political and spiritual purpose of freedom, in sharp contrast with “the business case;” their prescience remains unrecognized as a theoretical contribution to the field.

The creation stories of CRE, Indigenous evaluation, and other frames that interrogate and push against the accepted roots of evaluation are thus less about a mid-19th-century slave ship and more about the idea that those suffering under—and resisting—the ongoing effects of colonization, war, and enslavement are the experts of our own endeavors and determine our own fate, with the *kujichagulia* (Ngunzo Sabo notion of self-

determination) that our forebears intended (Battiste, 2000; Chilisa; 2012; Chilisa, Major, Gaotlhobogwe, & Mokgolodi, 2016; Kovach, 2010; LaFrance & Nichols, 2009; LaFrance & Nichols; 2010). As such, we mutiny our captors by challenging our bondage within systems of knowledge production that capitalize on difference and identity without materially shifting toward structures that are relational, reciprocal, and connected to the lived experiences of groups not coincidentally overrepresented among program participants. As students of history, we expect these acts of courage to be variously rewarded or punished by those who continue to exercise disproportionate control over our livelihood through government policies, development aid, funding proclamations and declarations, and accountability regimes. This includes quite literal disciplining from academic institutions and individuals who remain stifled by narrow ways of determining whose knowledge should be valued, cherished, and reproduced, and how. Seeing the CRE moniker on the third edition of *Evaluation Roots* is not the destination. Rather, liberation lies in defining, in our own languages, the structural relations undergirding our work—in the academy and in the worlds that we study and evaluate—for ourselves, based on our own logics and values, with and for the people with whom we are mutually accountable, without fear of punishment and at no one else's mercy. We end with a quote by Gordon and colleagues (1990) on the recognition of liberation as a goal of science, and likely of evaluation, as we authors reconsider the value and purpose of our lifework:

Liberation is a value worthy of science. That should be the perspective from which the minority scientist seeks to advance knowledge, always in the spirit of respect for logical canons, multiple perspectives, and methodological rigor; not for the purpose of simply predicting, controlling and understanding, but for the purpose of emancipating (liberating) the bodies, minds, communities, and spirits of

oppressed mankind. We join them in doing so, and invite you to do the same. (p. 19)

ACKNOWLEDGMENTS

Thanks to Dawn Adams, Katrina Bledsoe, Tina Christie, Jennifer Greene, Karen Kirkhart, and Stafford Hood for their reading, review, and critique to earlier versions of this chapter.

NOTE

1. Call to Action on Unearthing Evaluation Roots can be found at the following: https://docs.google.com/forms/d/e/1FAIpQLSeNojhYI2RWlIqaKXfGGe97m5mcCwrywXuedcaOJlw4_YNAgg/viewform

REFERENCES

- Battiste, M. (Ed.). (2000). *Reclaiming indigenous voice and vision*. University of British Columbia Press
- Carden, F., & Alkin, M. C. (2012). Evaluation roots: An international perspective. *Journal of MultiDisciplinary Evaluation*, 8(17), 102–118.
- Cavino, H. (2013). Across the colonial divide: Conversations about evaluation in Indigenous contexts. *American Journal of Evaluation*, 34(3), 339–355.
- Chilisa, B. (2012). *Indigenous research methodologies*. SAGE.
- Chilisa, B., Major, T. E., Gaotlhobogwe, M., & Mokgolodi, H. (2016). Decolonizing and indigenizing evaluation practice in Africa: Toward African relational evaluation approaches. *Canadian Journal of Program Evaluation*, 30(3), 313–328.
- Collins, P. H. (1998). It's all in the family: Intersections of gender, race and nation, *Hypatia*, 13(3), 62–82.
- Collins, P. H. (2008). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. (2nd ed.). Routledge.
- Crenshaw, K. W. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In K. T. Bartlett & R. Kennedy (Eds.), *Feminist legal theory: Readings in law and gender* (pp. 139–167). Routledge.

- Fine, M. (1998). Working the hyphens: Reinventing self and others in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (pp. 130–155). SAGE.
- Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). A guide to conducting culturally responsive evaluations. In J. Frechling (Ed.), *The 2010 user-friendly handbook for project evaluation* (pp. 75–96). National Science Foundation.
- Gordon, E. W., Miller, F., & Rollock, D. (1990, April). Coping with the communicentric bias in knowledge production in the social sciences. *Educational Researcher*, pp. 14–19.
- Hood, S. (1998). Responsive evaluation Amistad style: Perspectives of one African American evaluator. In R. Davis (Ed.), *Proceedings of the Stake Symposium on Educational Evaluation* (pp. 101–112). University of Illinois.
- Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. In J. C. Greene & T. A. Abma (Eds.), *Responsive evaluation. New Directions for Evaluation*, 92, 31–44.
- Hood, S. (2017). Continuing the exploration of African Americans in the early history of evaluation in the United States: Contributions of Ambrose Caliver in the U.S. Office of Education. *American Journal of Evaluation*, 38(2), 262–274.
- Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the “Nobody Knows My Name” Project, and African educational excellence. *Review of Educational Research*, 78(3), 410–426.
- Hood, S., & Hopson, R. (2018). Race and the production of knowledge in Black higher education: The legacy and contributions of Charles H. Thompson and the *Journal of Negro Education* in evaluation. In S. Graham, B. Foster, & J. Donaldson (Eds.), *Transformation of higher education: Anthropological and policy perspectives* (pp. 237–252). Information Age.
- Hood, S., Hopson, R., & Frierson, H. (Eds.). (2005). *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice*. Information Age.
- Hood, S., Hopson, R., & Frierson, H. (Eds.). (2015). *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. Information Age.
- Hopson, R. K. (2009). Reclaiming knowledge at the margins: Culturally responsive evaluation in the current evaluation moment. In K. Ryan & J. B. Cousins (Eds.), *The SAGE international handbook of educational evaluation* (pp. 429–446). SAGE.
- Hopson, R., & Hood, S. (2005). An untold story in evaluation roots: Reid E. Jackson and his contributions toward culturally responsive evaluation at three quarters of a century. In S. Hood, R. Hopson & H. Frierson (Eds.), *The role of culture and cultural context in evaluation: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 85–104). Information Age.
- Kirkhart, K. E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice*, 16(1), 1–12.
- Kovach, M. (2010). *Indigenous methodologies: Characteristics, conversations, and contexts*. University of Toronto Press.
- LaFrance, J., & Nichols, R. (2009). *Indigenous evaluation framework: Telling our story in our place and time*. American Indian Higher Education Consortium (AIHEC).
- LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an indigenous evaluation framework. *Canadian Journal of Program Evaluation*, 23(2), 13–31.
- Lopez, A., Rios, A., Scott, S., Shanker, V., Smith, L., & Camman, C. (2021, November 6). *Unearthing evaluation's roots: Decolonizing Evaluation Week*. <https://aea365.org/blog/decolonizing-evaluation-week-unearthing-evaluations-roots-by-carolyn-camman-albertina-lopez-aisha-rios-sheri-scott-vidhya-shanker-and-libby-smith>.
- Madison, A. M. (1992). Primary inclusion of culturally diverse minority program participants in the evaluation process. *New Directions for Program Evaluation*, 53, 35–43.
- Okun, T. (1999). *White supremacy culture*. <https://www.whitesupremacyculture.info/>
- Page, S. E. (2019). *The diversity bonus: How great teams pay off in the knowledge economy*. Princeton University Press.
- Patton, M. Q. (2017). *Principles-focused evaluation: The GUIDE*. Guilford Press.
- Peak, G. L., Peters, L., & Fishman, S. (2007). *Tracking transformation: Evaluating the American Evaluation Association's Building Diversity Initiative* (Report of Phase I Evaluation Findings). www.eval.org/p/cm/ld/fid=131

- Ray, V. (2019). A theory of racialized organizations. *American Sociological Review*, 84(1), 26–53.
- Scheurich, J. J., & Young, M. D. (1997). Coloring epistemologies: Are our research epistemologies racially biased? *Educational Researcher*, 26, 4–16.
- Schwandt, T. (2008). The relevance of practical knowledge traditions to evaluation practice. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 29–40). Guilford Press.
- Shanker, V. (2019). *Definitional tension: The construction of race in and through evaluation*. Unpublished doctoral thesis, University of Minnesota. <https://conservancy.umn.edu/handle/11299/211799>
- Shanker, V., & Maikuri, A. (2021). “They are intertwined”: Harm and the M&E cycle. https://wiredspace.wits.ac.za/bitstream/handle/10539/30481/harm_and_the_me_cycle.pdf?sequence=3&isallowed=y
- Stanfield, J. H., II (1985). The ethnocentric basis of social science knowledge production. *Review of Research in Education*, 12, 387–415.
- Stanfield, J. H., II (1999). Slipping through the front door: Relevant social scientific evaluation in the people of color century. *American Journal of Evaluation*, 20(3), 415–431.
- Stanfield, J. H., II (Ed.). (2011). *Rethinking race and ethnicity in research methods*. Left Coast Press.
- Stanfield, J. H., II, & Dennis, R. M. (Eds.). (1993). *Race and ethnicity in research methods*. Sage.
- Symonette, H. (2004). Walking pathways toward becoming a culturally competent evaluator: Boundaries, borderlands, and border crossings. *New Directions for Evaluation*, 102, 95–109.
- Symonette, H. (2015). Culturally responsive evaluation as a resource for helpful-help. In S. Hood, R. Hopson, & H. Frierson, (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 109–130). Information Age.
- Ward Hood, D., & Cassaro, D. A. (2002). Feminist evaluation and the inclusion of difference. *New Directions for Evaluation*, 96, 27–40.
- Weisinger, J. Y., Borges-Méndez, R., & Milofsky, C. (2015). Diversity in the nonprofit and voluntary sector. *Nonprofit and Voluntary Sector Quarterly*, 45(1, Suppl.), 3S–27S.
- Zuberi, T., & Bonilla-Silva, E. (Eds.). (2008). *White logic, white methods: Racism and methodology*. Rowman & Littlefield.

CHAPTER 12

A Transformative Lens in Evaluation A Pathway to Transformative Change

Donna M. Mertens

INTRODUCTION

Increasing disparities in wealth between rich and poor, natural disasters, violence against people of color and mass incarceration, immigration crises, violations of the rights of women and people with disabilities, and a global pandemic grab our attention and leave us with the question: What can we do as a community of evaluators? The global community has recognized the need for transformative change if the goals of reduced poverty and hunger, improved health and education, increased equity and access to clean water and energy, employment, worldwide peace, and a clean environment are to be achieved (van den Berg, Magro, & Mulder, 2019). The Sustainable Development Goals (SDGs) that were adopted by the United Nations (UN General Assembly, 2015) and endorsed by the global community encompass these concerns. Although the United Nations set a time frame for achievement of the SDGs by 2030, the problems being addressed are

entrenched in a system that is not likely to yield definitive solutions in the time allotted. Thus, evaluators have a role to play in supporting the achievement of the SDGs through 2030 and beyond. In this chapter, I argue that the use of a transformative theoretical lens enables evaluators to contribute to the achievement of these SDGs and to provide a platform for future progress toward a more equitable world. The goal of transformative evaluation is to support the development of culturally responsive interventions that foster increased respect for human rights and achievement of social, economic, and environmental justice (Mertens, 2020b, 2018; Mertens & Wilson, 2019).

ASSUMPTIONS OF THE TRANSFORMATIVE PARADIGM

The salient values that are associated with a transformative stance include being culturally responsive, explicitly addressing inequities, incorporating reciprocity, and

recognizing community resilience and the interconnectedness of all things (human and nonhuman). The ethical assumptions associated with a transformative stance inform decisions about methodology. For example, if we work from the assumption that our evaluations need to further human rights and increase social, economic, and environmental justice, then how does that affect our methodological choices? These values lead to decisions to structure the study to be inclusive of all voices and use strategies that support the development of interventions that are responsive and supportive of positive transformative change.

Following from the transformative ethical assumptions, evaluators recognize that there are different versions of reality informing the understanding of problems and development of interventions. Evaluators have an ethical responsibility to bring to visibility and challenge versions of reality that sustain an oppressive status quo and to reveal and support those that provide a pathway toward enhanced justice and human rights. For example, decades of data collection in the United States indicate that chronic health disparities exist between White people and Black and Indigenous people and People of Color (BIPOC) that are manifest in higher mortality rates and lower overall well-being in BIPOC communities (Estrella, Inouye, & Lewis-Chapman, 2019). As the world experiences a pandemic of epic proportions, the consequences of these disparities are magnified in that members of BIPOC communities are more likely to be infected by and die from COVID-19 than are White people (Adhikari et al., 2020). Evaluators can play a role in advancing knowledge about the conditions that lead to the increased infections and deaths, and to the development of transformative interventions (Mertens, 2020a).

One version of reality that is used to explain health disparities is that members of BIPOC communities need to make better choices to eat healthier foods and exercise. However, this version of reality ignores

the conditions created by systemic racism that contribute to health disparities, such as poverty, unemployment, educational access, wealth, access to grocery stores and healthy food, safe places to exercise, stress associated with poverty and systemic discrimination, residential segregation, and access to affordable and culturally responsive health care (Mertens, 2020a). Use of a transformative lens would lead an evaluator to recognize that basing an intervention on a version of reality that attributed the causes to bad choices in food and lack of exercise would continue to support an oppressive status quo. However, structuring an evaluation to be inclusive of all voices, valuing the knowledge and experiences of members of BIPOC and other marginalized communities, consciously addressing power issues, and providing safe conditions for collection of data can contribute to a better and more nuanced understanding of community needs in order to develop an intervention that has greater potential to decrease health disparities.

The inclusion of all voices is replete with issues of power differences and has implications for evaluation methods. These power differences can be addressed in multiple ways. First, transformative evaluators would strive to be inclusive in culturally responsive ways that place value on the experiences of members of marginalized communities (Mertens & Wilson, 2019). This means taking the time to form relationships of trust and figuring out working strategies that are acceptable and feasible for all stakeholders. Strategies for inclusion might include the use of advisory boards or having community members as co-evaluators. Capacity building to enhance stakeholders' skills in evaluation and in strategies of social advocacy can contribute to supporting continued actions that would be necessary to sustain transformative change after the conclusion of an evaluation study.

In the transformative spirit, evaluators might take a page from the work of social activists who have organized groups to cre-

ate needed change. One example of social activism is found in the organizing of fast-food workers to increase their minimum wage, known as Fight for \$15 (Greenwood, 2019). The Service Employees International Union partnered with a grassroots group called New York Communities for Change. They started a coalition of fast-food workers through hiring local organizers to visit apartments, houses, and fast-food restaurants. They initially held meetings in New York City, where workers could voice their experiences and desires. The workers decided they wanted to go on strike, even though they were not union members. The first one-day strike occurred at 27 locations in New York City, including the McDonald's that is situated on Madison Avenue and 40th Street and a Wendy's on 34th Street north of Madison Square Garden and Penn Station. These locations were advantageous in that large numbers of people quickly became aware of the strike; individuals tweeted and posted on social media; and local and national media provided extensive coverage. Political leaders took note and put pressure on businesses to raise wages. This movement spread across the country, and 20 states passed a minimum wage law by November 2021. Even though this movement occurred outside of the evaluation world, it demonstrates the power of a coalition for taking action and sustaining change.

The idea of coalitions in evaluation work is not a foreign concept. Wolfe, Price, and Brown (2020) note that community coalitions have become the norm in programs that address public health and social problems. They are required by some of the major funding agencies, such as the Health Resources and Services Administration Healthy Start Program, the Centers for Disease Control and Prevention cancer control programs, and the Substance Abuse and Mental Health Services Administration Drug Free Communities. Wolfe and colleagues' work focuses on evaluating the coalitions themselves. The transformative lens would push the role of coalitions fur-

ther to ask how the coalitions could become an integral part of the evaluation planning, implementation, and use.

Another contribution of a transformative lens in evaluation rests on its goal of reducing or eliminating discrimination and oppression (Mertens & Wilson, 2019). The dimensions of diversity that are relevant are contextually dependent. Many of the examples already discussed are based on racial/ethnic discrimination in the United States. However, discrimination can be based on multiple characteristics, such as gender, sexual identity, disability, religion, geographic area, socioeconomic status, Indigeneity, Deafness, and age. These dimensions of diversity might be combined in specific contexts to address the effects of intersectionality. There are also contexts in which concern with BIPOC communities as conceptualized in the United States would be less relevant. For example, when conducting evaluation in low- and middle-income countries, gender and disability are recognized bases for discrimination (UN Women, 2017; United Nations, 2018). (It should be noted that discrimination against women and girls is also a problem in higher-income countries such as the United States [Conrad, 2020].) Violence against women has long been a priority for UN Women, but the urgency for addressing this issue is amplified by the increase in violence against women across the globe during the COVID-19 pandemic (Mlambo-Ngcuka, 2020). Being a member of a sexual minority group (lesbian, gay, bisexual, transgender, queer, and intersex [LGBTQI]) is another basis of discrimination that is made worse by the criminalization of same-sex behavior in several countries (e.g., Zimbabwe, Cameroon, Uganda; Miller, 2020). Use of a transformative lens guides evaluators to conduct a contextual analysis in order to understand the dimensions of diversity that are relevant within a specific context.

The preceding discussion of ethics from a transformative perspective has implications for methodological choices and evaluation designs. The need to establish culturally

responsive relationships with the full range of stakeholders and the conduct of a contextual analysis have already been raised. These initial stages of a transformative evaluation lay the groundwork for informing a more nuanced understanding of problems, strengths, and cultural and contextual complexities relevant to the evaluation. They provide an evidence base that can guide what is needed in terms of interventions that are culturally responsive and sustainable.

Transformative evaluators are quite eclectic in their choices of methodologies; however, it is likely that a mixed methods design would be chosen in order to be responsive to the full range of stakeholders needs (Mertens, 2018). Transformative mixed-methods designs can include case studies, randomized controlled trials, participatory action research, community-based designs, oral histories, surveys, focus groups, and many other options. What is unique about this theoretical lens is that the design would be interrogated to ensure that the voices of all stakeholders are included, power relations are critically examined, methods and interventions would be culturally responsive, and a focus on supporting transformative action. Before providing specific examples to illustrate these applications, the common ground between transformative, equitable, culturally responsive, and Indigenous lenses for evaluation is discussed.

POROUS BORDERS AND CLOSE RELATIONS: CULTURALLY RESPONSIVE, INDIGENOUS, AND EQUITABLE EVALUATION

The transformative lens aligns with other approaches that prioritize issues of justice and transformative change. Culturally responsive evaluation (CRE) is one such approach that is rooted in critical race theory and had its beginnings with a small group of African American evaluators (Hood, Hopson, & Frierson, 2005). Thus, culturally responsive evaluation is rooted

in concerns with culture, history, and context in diverse ethnic, linguistic, and racial communities of color (Hood, Hopson, & Frierson, 2015). BIPOC communities are well represented in the application of the CRE approach (Chouinard & Cram, 2020). The application of a CRE approach in marginalized communities on the basis of other dimensions of diversity beyond race is considered to be new territory for this approach (Hood et al., 2015).

Equitable evaluation (Dean-Coffey, Casey, & Caldwell, 2014; Equitable Evaluation Initiative, 2020) is another approach that aligns with a transformative lens and is characterized by a focus on advancing equity, critically interrogating the effect of history and structure, and manifesting multicultural validity and participant ownership. Equitable evaluation has been combined with culturally responsive evaluation to create an approach known as culturally responsive equitable evaluation (CREE), defined as follows:

Culturally responsive and equitable evaluation requires the integration of diversity, inclusion, and equity in all phases of evaluation. CREE incorporates cultural, structural, and contextual factors (e.g. historical, social, economic, racial, ethnic, gender) using a participatory process that shifts power to individuals most impacted. CREE is not just one method of evaluation, it is an approach that should be infused into all evaluation methodologies. CREE advances equity by informing strategy, program improvement, decision-making, policy formation, and change. (Change Matrix, 2020)

Indigenous scholars have explored the boundaries between Indigenous approaches and transformative CREs in their communities (Chilisa & Mertens, 2021; Cram & Mertens, 2015). Indigenous evaluators share the goal of achieving justice in their communities and the importance of developing respectful relationships. However, there are important differences because of their his-

tory of colonization and their legal status as sovereign nations (in some parts of the world). They also recognize the spiritual connections that connect us all; this is not a concept typically included in evaluation approaches from the higher-income countries.

The porous borders between culturally responsive, equitable, and Indigenous theoretical perspectives means that evaluators can and do combine these lenses in their work in the pursuit of increased justice.

SCOPE: MULTIDISCIPLINARY, LOCAL AND GLOBAL, AND PROJECT, PROGRAM, SYSTEMS, AND POLICY

Applications of a transformative lens in evaluation are appropriate in any context in which inequities are present that impede access for all to social, economic, and environmental justice. Examples of detailed application of the transformative lens are presented later in this chapter to illustrate specific implications of using this lens. The breadth of its applicability is demonstrated by the number of sectors in which it has been applied, the location of the applications, the type of the evaluand, the populations included in such evaluations, and the level of the evaluand, as seen in the following examples:

- *Education:* Barnhardt, Reyes, Rodriguez, and Ramos (2018) evaluated educational policies related to inclusion of undocumented students in higher education in the United States. Flecha (2014) used a transformative lens to evaluate systemic changes in education for the Roma population in Spain.
- *Mental health:* Liebenberg, Jamal, and Ikeda (2020) applied it to an evaluation of a single project in an Indigenous village in Canada designed to strengthen mental health.
- *Environmental policy:* Beauchamp, Hoff-

mann, D'Errico, Lucks, and El-Saddik (2020) advocate for a transformative lens to inform climate change policies across the globe in evaluations related to achieving the SDGs to achieve action on climate change.

- *Criminal justice systems:* Todrys, Amon, Malembeka, and Clayton (2011) evaluated policies and legislation related to the criminal justice system in Zambia. Drill, Stack, Qureshi, and Lahoff (2017) evaluated a transition program for youth involved in the criminal justice system in Oregon.
- *Refugees:* Habashi and Worley (2009) used a transformative approach to evaluate programs for Palestinians in refugee camps.
- *Economic development:* Widianingsih and Mertens (2019) applied a transformative lens to economic development policies and programs in Bandung, West Java, with attention to the intersection of economic, social, and environmental justice.
- *Gender-based violence:* Puigvert, Valls, Yeste, Aguilar, and Merrill (2019) used a transformative lens for evaluation of a systemwide change at Spanish universities to eliminate gender-based violence.
- *Agriculture:* Arko-Achemfuor, Romm, and Serolong (2019) applied a transformative lens to the evaluation of a program for farmers in South Africa.

As can be seen, the transformative lens has been applied locally and globally. Many different communities have found the use of a transformative lens to be useful (e.g., people with disabilities, Deaf people, women and girls, racial and ethnic minorities, farmers in low- and middle-income countries, sexual minorities, refugees and immigrants, and Indigenous communities). I describe in the next section the specific characteristics of two studies that illustrate the use of a transformative lens the health and impact investing sectors.

EVALUATIONS WITH A TRANSFORMATIVE LENS

Many examples of evaluations that used a transformative theoretical lens were listed in the previous section; more are available in the literature (see Mertens, 2018, 2020b; Mertens & Wilson, 2019). The specific examples in this section were chosen to present methodological implications in more detail when the transformative lens is applied in practice. The first example involves the LGBTQI community, which faces stigma, discrimination, and violence while trying to access HIV health care in many African and Caribbean countries (Miller, 2020). The Advocacy and other Community Tactics Project (Project ACT) was undertaken to evaluate a collaboration among a lead activist agency, a watchdogging expert, and nine collaborating identity groups that set out as a collective to document the quality of health care for gay/bi men and trans women, so that they could push these local facilities to change. At first glance, the actual evaluation data collection looks rather traditional. "Evaluation site visits incorporated observations, interviews, and focus groups, and review of documents" (Miller, 2020, p. 1). However, the transformative aspects of the evaluation are revealed by the advocacy stance taken by the collaboration on behalf of an oppressed group, ensuring grassroots representation, providing a supportive environment, focusing on justice, and consciously addressing power issues to form a basis for transformative action. The collaboration guided decisions about data collection by members of the LGBTQI communities and provided a mechanism for sustainability.

A contextual analysis revealed legislative barriers in the form of the criminalization of same-sex relations, with a prison sentence up to 5 years. The evaluator also documented the incidence of homophobic and transphobic violence in the country through the use of extant data and interviews with activists. These data revealed oppressive cultural practices such as parents rejecting

their LGBTQI children and landlords refusing to rent to or evicting members of this community. Local data collectors chosen by the identity organizations attended training sessions on what to do when visiting the health service agencies. They used a standardized tool to record their data that had been developed in collaboration with the identity organizations, so that it reflected the indicators of importance to that community.

The evaluation documented that the local data collectors needed better safety protections while going into these settings, better training to go into these settings successfully, more funds, and basic social supports to do the dangerous and stressful work they were undertaking. Several modifications were made; for example, the local data collectors encountered hostility to the degree that they asked for additional training on how to behave in a nonprovocative manner when experiencing discrimination. The local data collectors also asked for money to pay for the medical fees and transportation costs; this was then provided by the project. Four of the nine identity organizations contributed to the sustainability of the project by training additional people to collect data. The project also introduced support meetings for the data collectors, so they could complete their work while dealing with the stress they experienced during fieldwork.

The evaluation also documented the transformative uses based on the locally collected data, such as training health care workers and district directors. As a result, facilities and districts developed corrective action plans, committing to ongoing training to reduce discriminatory practices, formation of a WhatsApp group for district directors to stay informed as issues arose, and establishment of a health care site at one of the identity organizations. In addition, prison health care workers asked for training for their entire staff. The local data collectors reported they were more likely to seek health care and to help others navi-

gate the system. While the project was not focused on changing the law in Cameroon, it did open up access to a more responsive health care system for this marginalized community.

The second example is taken from the investment impact community with a project designed to improve educational outcomes for students living in poverty in Kenya (Bolinson & Mertens, 2021; Wakiaga & Bolinson, 2019). Engineers without Borders Canada funded M-Shule, a seed-stage enterprise that used a mobile education platform to engage students in upper primary school in Kenya. They commissioned a transformative evaluation during the postinvestment stage, meaning after the enterprise had been in operation for approximately a year and was ready to assess its current status and determine future actions.

The evaluation included several phases that exemplify application of a transformative lens. First, local evaluators were hired through a collaborative effort with the African Evaluation Association. The evaluators formed a relationship with the M-Shule staff and established a local advisory committee (LAC) made up of a representative group of beneficiaries. The selection of LAC members involved making contact with local school associations, attending meetings at the schools, and meeting the head teachers, who invited the evaluators to visit the schools. The LAC included head teachers, school director, parents, teachers, and students. The first meeting was held before the evaluation plan was finalized to obtain input on the evaluation questions, methods used, sampling, recruitment, and logistics. For example, one question that emerged from this process was “How are the cultural norms and beliefs of the intended beneficiaries incorporated into M-Shule?” (Wakiaga & Bolinson, 2019, p. 11)

Conduct of a contextual analysis provided insights into the social, political, and economic context, and provided data on Kenyan education policy, education enrollment and achievement, and trends in the use of

digital technology. Data collection methods included focus groups, key informant interviews, and short surveys; the entry into the schools and communities was informed by recommendations from the LAC. The language used was either Kiswahili or English, depending on the preference of the stakeholders. Another transformative principle was incorporated into the data sharing and interpretation phase. The LAC designed the session for sharing and co-interpretation of the results to support reciprocity and ownership of the results. Children were separated from adults at this stage, as earlier feedback indicated that the children did not feel they could express themselves freely in front of the adults. The findings were used for transformative purposes by M-Shule, the school association, and the broader impact investment community. For example, M-Shule forged stronger relationships with the user community and revised their fee structure for the service, obtaining funding through foundations for the families that could not afford the small fee. The school association and head teachers indicated that they wanted to engage in additional collaborations with M-Shule and integrate the educational resources into their instruction. Three products were shared with the impact investment community: an evaluation report (Wakiaga & Bolinson, 2019), a toolkit for conducting transformative evaluations in the impact investing sector (Bolinson & Mertens, 2019), and a chapter in a handbook on impact investing (Bolinson & Mertens, 2021).

Identifying a Transformative Theoretical Approach: How Do I Know When I See It?

Several scholars have identified what they consider to be the key characteristics of an evaluation that exemplify the application of a transformative theoretical lens (Bolinson & Mertens, 2021; Canales, 2013; Mertens & Wilson, 2019; Sweetman et al., 2010). Evaluations that are informed by a transformative lens can take many forms; there is no rigid

set of criteria that define this approach. Key characteristics include the following:

- Begins with an ethical stance that is inclusive and culturally responsive, consciously addresses inequities, provides for reciprocity, recognizes community resilience and strengths, and focuses on supporting transformative change in the form of increased social, economic, and environmental justice.
- Supports the development of collaborations and coalitions that are informed by the principles of social activism as a way to generate ownership and sustainability.
- Does not make assumptions about what the problems and solutions are before consultation with the full range of stakeholders.
- Makes visible versions of reality that support an oppressive status quo and reveals versions of reality that can lead to increased justice.
- Commits to recognizing heterogeneity and intersectionality within communities based on relevant dimensions of diversity that are used as a basis of discrimination.
- Incorporates culturally responsive, equity-focused, anti-racist, feminist, and Indigenous approaches as appropriate to context.
- Uses a participatory and reflective entry process into a community that is designed to build trust, address power differences, and give value to grassroots knowledge.
- Employs strategies throughout the evaluation that build capacity of the stakeholders to use the findings for transformative purposes and disseminates findings to support increased social, economic, and environmental justice.
- Has an action orientation; the structure, process and findings of the evaluation are designed to support transformative change.
- Does not overpromise to stakeholders

that transformative changes will occur as a result of their participation in the evaluation, but commits to structuring the evaluation in ways that increase the probability of such change.

A transformative-informed evaluation may not include all of these elements, but they should be planned, conducted, and used with a transformative spirit that not only does no harm but also contributes to the greater good.

CONCLUSION

The global community is conscious of the need for transformative change to address intransigent problems associated with inequities and injustice. The use of a transformative lens in evaluation provides one pathway to inform understandings that underlie the development of culturally responsive interventions. Use of this lens is widespread, even with the many challenges that are inherent in adopting this stance. It is not easy to include the voices of all stakeholders; yet failing to do so has inhibited the global communities' ability to make the transformative changes that are necessary. Evaluators can work together to overcome these challenges in order to be meaningful partners in this push for increased justice. The key is centering economic, environmental, and social justice in evaluation theory and practice.

REFERENCES

- Adhikari, S., Pantaleo, N. P., Feldman, M. M., Ogedegbe, O., Thorpe, L., & Troxel, A. B. (2020). Assessment of community-level disparities in coronavirus disease 2019 (COVID-19): Infections and deaths in large US metropolitan areas. *JAMA Network Open*, 3(7), 1–4.
- Arko-Achemfuor, A., Romm, N., & Serolong, L. (2019). Academic–practitioner collaboration with communities towards social and ecological transformation. *International Journal of Transformative Research*, 6(1), 1–9.

- Barnhardt, C. L., Reyes, K., Rodriguez, A. V., & Ramos, M. (2018). A transformative mixed methods assessment of educational access and opportunity for undocumented college students in the Southeastern United States. *Journal of Mixed Methods Research*, 12(4), 413–436.
- Beauchamp, E., Hoffmann, D., D'Errico, S., Lucks, D., & El-Saddik, K. (2020, September). The climate crisis: Four ways evaluations can strengthen policymaking. International Institute for Environment and Development Briefing.
- Bolinson, C., & Mertens, D. (2019). *A transformative evaluation toolkit for the impact investing sector: How entrepreneurs and impact investors can deepen their data collection*. Engineers Without Borders Canada.
- Bolinson, C., & Mertens, D. M. (2021). Transformative evaluation and impact investing: A fruitful marriage. In R. P. Herman & E. de Moraes Sarmiento (Eds.), *Global handbook of impact investing* (pp. 697–734). Wiley.
- Canales, G. (2013). Transformative, mixed methods checklist for psychological research with Mexican Americans. *Journal of Mixed Methods Research*, 7(1), 6–21.
- Change Matrix. (2020). *Expanding the Bench: Understanding the terms we use*. <https://expandingthebench.org/about-etb/understanding-terms-we-use>
- Chilisa, B., & Mertens, D. M. (2021). Indigenous Made in Africa evaluation frameworks: Addressing epistemic violence and contributing to social transformation. *American Journal of Evaluation*, 42, 241–253.
- Chouinard, J. A., & Cram, F. (2020). *Culturally responsive approaches to evaluation*. SAGE.
- Conrad, J. A. (2020). On intellectual and developmental disabilities in the United States: A historical perspective. *Journal of Intellectual Disabilities*, 24(1), 85–101.
- Cram, F., & Mertens, D. M. (2015). Transformative and indigenous frameworks for multi-method and mixed methods research. In S. N. Hesse-Biber & B. B. Johnson (Eds.), *The Oxford handbook of mixed and multi methods research* (pp. 99–109). Oxford University Press.
- Dean-Coffey, J., Casey, J., & Caldwell, L. D. (2014). Raising the bar—integrating cultural competence and equity: Equitable evaluation. *Foundation Review*, 6(2).
- Drill, K., Stack, E. E., Qureshi, C., & Lahoff, R. (2017). *Oregon's African American/Black student success plan*. Oregon Department of Education Office of Equity, Diversity and Inclusion.
- Equitable Evaluation Initiative. (2020). *The Equitable Evaluation Framework Principles and Co-Learning*. www.equitableeval.org/ee-framework
- Estrella, R., Inouye, T. E., & Lewis-Chapman, H. (2019). *Centering race in health equity advocacy*. The Colorado Trust.
- Flecha, R. (2014). Using mixed methods from a communicative orientation: Researching with grassroots Roma. *Journal of Mixed Methods Research*, 8(3), 245–254.
- Greenwood, S. (2019). *Beaten down, worked up*. Knopf.
- Habashi, J., & Worley, J. (2009). Child geopolitical agency. *Journal of Mixed Methods Research*, 3(1), 42–64.
- Hood, S., Hopson, R., & Frierson, H. (Eds.). (2005). *The role of culture and cultural context*. Information Age.
- Hood, S., Hopson, R., & Frierson, H. (2015). This is where we continue to stand. In *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. (pp. ix–xvii). Information Age.
- Liebenberg, L., Jamal, A., & Ikeda, J. (2020). Extending youth voices in a participatory thematic analysis approach. *International Journal of Qualitative Methods*, 19, 1–13.
- Mertens, D. M. (2018). *Mixed methods design in evaluation*. SAGE.
- Mertens, D. M. (2020a, March). *Ethical implications of transformative mixed methods research and evaluation*. Presentation at the Robert Wood Johnson Foundation 2020 Sharing Knowledge: Racial Justice and Health Conference, Jackson, MS.
- Mertens, D. M. (2020b). *Research and evaluation in education and psychology* (5th ed.). SAGE.
- Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice* (2nd ed.). Guilford Press.
- Mlambo-Ngcuka, P. (2020). *Violence against women and girls: The shadow pandemic*. UN Women.
- Miller, R. L. (2020). *Reducing stigma and discrimination in access to HIV health care for gay and bisexual men and transgender women using mystery patients in Cameroon and Zimbabwe*. Michigan State University.
- Puigvert, L., Valls, R., Yeste, C., Aguilar, C., & Merrill, B. (2019). Resistance to and transformations of gender-based violence in Span-

- ish universities: A communicative evaluation of social impact. *Journal of Mixed Methods Research*, 13(3), 361–380.
- Sweetman, D., Badiie, M., & Creswell, J. (2010). Use of the transformative framework in mixed methods studies. *Qualitative Inquiry*, 16(6), 441–454.
- Todrys, K. W., Amon, J. J., Malembeka, G., & Clayton M. (2011). Imprisoned and imperiled: Access to HIV and TB prevention and treatment and denial of human rights in Zambian Prisons. *Journal of the International AIDS Society*, 14(8), 1–11.
- UN General Assembly. (2015). *Resolution adopted by the General Assembly on 25 September 2015. Transformative our world: The 2030 agenda for sustainable development*. Author.
- UN Women. (2017). *Making the SDGS count for women and girls with disabilities*. Author.
- United Nations. (2018). *UN Flagship Report on Disability and Development: Realization of the sustainable development goals by, for and with persons with disabilities: Disability and the 2030 Agenda for Sustainable Development*. <https://social.un.org/publications/un-flagship-report-disability-final.pdf>
- Van den Berg, R. D., Magro, C., & Mulder, S. S. (Eds.). (2019). *Evaluation for transformational change*. IDEAS.
- Wakiaga, L., & Bolinson, C. (2019). *Using transformative evaluation in impact investing: A case study with Engineers Without Borders Canada and M-Shule*. Engineers Without Borders Canada.
- Widianingsih, I., & Mertens, D. M. (2019). Transformative research and the sustainable development goals: Challenges and a vision from Bandung, West Java. *International Journal of Transformative Research*, 6(1), 27–35.
- Wolfe, S. M., Price, A. W., & Brown, K. K. (Eds.). (2020). Evaluating community coalitions and collaboratives. *New Directions for Evaluation*, 165, 1–185.

CHAPTER 13

Culturally Responsive Indigenous Evaluation

An Indigenous Journey through Culturally Responsive Approaches

Fiona Cram
Jill Anne Chouinard

INTRODUCTION

Our understanding of our place in this world and how we experience life is informed by our culture (Hopson, 2003). This includes the culture(s) of the people(s) we call our own, as well as how these people are positioned within the power structures of the culture that shapes our wider society (Sense, 2005; SenGupta, Hopson, & Thompson-Robinson, 2004). When we think about culture, we often understand that minoritized groups among us or people in other countries have it. It is often more difficult to see the unnamed culture of Whiteness that pervades Western societies, although this culture too is becoming more opaque as the privilege of Whiteness within racist societal culture continues to be unmasked. Culture, and the implications of culture for people's lived experiences, is everywhere.

For Indigenous peoples, these implications are both tangible and intangible. The

redistribution of resources, including land, into the hands of newcomers to their territories is a very tangible loss for many Indigenous peoples. This key source of Indigenous marginalization is felt in intangible or spiritual outcomes related to a loss of connectedness to and the disruption of guardianship responsibilities for the land. Whether under the flag of colonization or the mantra of development, outsiders continue to this day to enforce their notions of civilization and superiority to the detriment of the vitality of the cultures—knowledges, values, ways of life—of Indigenous peoples.

Nobles (1991) describes the similarities between political colonialism and scientific colonialism to highlight how scientists have often been staunch allies of colonial agendas (see Table 13.1, unshaded columns). Where colonial powers acquired Indigenous land, Nobles writes that the work of scientists provided a rationale for this theft. Likewise, Linda Smith (1992, p. 7) describes

the researchers who justify colonization as “willing bedfellows of assimilationist, victim-blaming policies” that dehumanize and marginalize Indigenous peoples. Indigenous peoples have been literally researched to death and, as George Sefa Dei (2006, p. 1) argues,

colonialism has not ended and we see around us today various examples of colonial and neo-colonial relations produced within our schools, colleges, universities, homes, families, workplaces and other institutional settings. It

is often said that globalization is the new word for imperialism.

Turning a corner on these practices in the pursuit of decolonization and the return of Indigenous territory, is a task to which scientists can also contribute. This can be a form of penance for the wrongs wrought by their disciplinary “ancestors,” as well as undertaken out of a commitment to being good partners and accomplices of Indigenous peoples in the sharing of this planet. Historians, for example, now acknowledge

TABLE 13.1. Comparative Colonialisms

	Site	Political colonialism	Scientific colonialism	Indigenous sovereignty
External Colonialisms	Access to Indigenous resources and knowledge	Colonial Power's assertion of their right to access, use, and benefit from the goods and resources of Indigenous peoples.	Scientist's assertion of their right to access, use, and benefit from knowledge sourced from Indigenous peoples.	Access to Indigenous people's communities and context is by invitation only and meetings follow local, indigenous protocols.
	Indigenous land	Colonial Power insists on settler acquisition of Indigenous land. Land is remade into property, a resource. Colonial Power is willing to go to war to uphold its “right” to the land.	Scientist endorses the contestation and consequent disruption of Indigenous relationships with the land—“a profound epistemic, ontological, cosmological violence” (Tuck & Yang, 2012, p. 5)	Indigenous communities regain, retain, or strengthen their place on and their relationships with their lands.
	Guardianship of Indigenous wealth and resources	The goods and resources of Indigenous peoples are exported by Colonial Powers, to be processed into manufactured products and/or wealth.	Knowledge gained from Indigenous peoples is exported by Scientists, to be processed into scientific products (i.e., books, articles, patents) and/or wealth.	Indigenous peoples maintain the guardianship over their resources, including their knowledge, and use these for the betterment of their people.
	Location of Indigenous resources and knowledge	Colonial Power is located outside Indigenous people's communities, so that colonial power and control are exerted from a distance.	Scientific knowledge and information gathered from and about Indigenous peoples is located outside Indigenous people's communities.	The sites of Indigenous resources and knowledge remain with Indigenous peoples, in their communities and are a source of their vitality and sustainability.
Internal Colonialisms	Control within “domestic” borders	“Biopolitical and geopolitical management of people, land, flora and fauna” (Tuck & Yang, 2012, p. 4)	Scientific evidence and rationalizations for a status quo of Indigenous inequality and disparity. Internalization of Indigenous deficits by population, including indigenous peoples.	Responsibility for the investigation and representation of Indigenous realities rests with Indigenous peoples.
	Authentication of “insiders”	Restrictive Indigenous registries designed to extinguish Indigenous claims to land.	Science as a tool to displace and supplant Indigenous epistemologies.	Indigenous decision making about who is Indigenous. Non-Indigenous evaluators “are beholden to the Indigenous laws and epistemologies of the lands” and territories they work in (Tuck & Yang, 2012, p. 6).

Note. Based on Nobles (1991, p. 296, Table 1).

that the burden of injustices and inequalities borne by Indigenous peoples are due to unnatural colonial events rather than any deficiency of Indigenous peoples (Anderson, 2007). It is time for evaluation to also turn this corner and challenge political and scientific colonialism through evaluation approaches that acknowledge and honor Indigenous sovereignty (Table 13.1, shaded column).

In this chapter on culturally responsive Indigenous evaluation, we therefore ask the same question as an elder speaking during the 1992 Canadian Royal Commission on Aboriginal Peoples: If Indigenous peoples have been researched to death, isn't it time we researched ourselves back to life (Brant Castellano, 2004). In other words, how can evaluation be a force in support of Indigenous cultural vitality, wellness, and sustainability? We begin with a short précis of the emergence of culturally responsive evaluation and its inclusion of Indigenous agendas. We then describe 10 dimensions that depict culture as part of the fabric of evaluation methodology, with a focus on Indigenous evaluation. Finally, we offer concluding remarks by way of "take home" messages about the important contribution that can be made to Indigenous lives by culturally responsive Indigenous evaluation and evaluators. Through the interconnectedness of Indigenous peoples and the land, we also contend that culturally responsive Indigenous evaluation is not a "branch," but rather the roots that secure the tree within the richness of the territory of evaluation.

A BRIEF HISTORY

At the headwaters of culturally responsive evaluation is a small group of African American researchers and evaluators—Reid E. Jackson, Leander Boykin, and Aaron Brown—who implemented culturally responsive evaluation methods during times of racial segregation from the 1930s to the 1950s (Hood, 2001). They sought to

expand evaluation beyond technical considerations because of their own beliefs in equality, democracy, and justice (Hood & Hopson, 2008). In the 1960s, a tributary of "responsive evaluation" was promoted by Robert Stake (1975) as he brought cultural diversity into the center of evaluation. The cultural dimensions of evaluation practice were further explored in the mid-1980s and 1990s, when our evaluation colleagues questioned "traditional" modes of evaluation, to highlight their incompatibility with issues of culture, intersectionality, and social justice (Kirkhart, 2005; MacDonald & Parlett, 1973; Mertens, 2009; Patton, 2002).

The role of culture came further to the fore in Stafford Hood's 1998 paper at the Robert Stake Retirement Symposium, and in 2012 he led the establishment of the Center for Culturally Responsive Evaluation and Assessment (CREA) at the University of Illinois (Frierson, Hood, & Hughes, 2002). Central to much of the discussion about culturally responsive evaluation is the importance of evaluators having a "shared lived experience with stakeholders" (Frierson et al., 2002, p. 84). This can be seen to be as much about the people themselves as it is about the places they occupy and the spaces (historical and contemporary) they inhabit within society. As this river of culturally responsive evaluation widened and deepened across the decades, we might imagine it being joined and strengthened by tributaries that strengthen our understanding of how people's "stories-so-far" (Massey, 2005, p. 9) have been defined by their cultural journey over a societal landscape.

From the late 1980s, Indigenous voices were also being heard more loudly within evaluation forums, through the work of Joan LaFrance and Richard Nichols (2009, 2010) in the United States and others around the world (e.g., Kawakami, Aton, Cram, Lai, & Porima, 2007). Culturally responsive approaches to evaluation strive to respond to calls from Indigenous peoples for the decolonization of evaluation (Smith, 2012) and for evaluation to contrib-

ute to a larger decolonization project that is about the return of Indigenous territories (Tuck & Yang, 2012). To achieve this, evaluators should understand the history of the Indigenous people they work with, alongside their needs, priorities, and aspirations, and set their evaluation within a relationship context that supports Indigenous self-determination and prosperity. In addition to promises contained in treaties made with newcomers, the right of Indigenous people to this reframing of evaluation is reiterated in the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). Part 1 of Article 31 of UNDRIP states:

Indigenous peoples have the right to maintain, control, protect and develop their cultural heritage, traditional knowledge and traditional cultural expressions, as well as the manifestations of their sciences, technologies and cultures. . . . They also have the right to maintain, control, protect and develop their intellectual property over such cultural heritage, traditional knowledge, and traditional cultural expressions.

LOCATING INDIGENEITY WITHIN CULTURALLY RESPONSIVE EVALUATION PRACTICE

Being responsive to culture and cultural context requires an understanding of the nexus of culture and evaluation, of how culture influences and shapes evaluation processes and consequences in local and global settings. In earlier research (see Chouinard & Cousins, 2009; Chouinard & Hopson, 2016), we identified seven dimensions where culture is dynamically implicated in evaluation (i.e., epistemological, ecological, methodological, political, personal, relational, institutional). Discussions between Cram and Mertens (2015, 2016) about the responsiveness of the transformative paradigm to the cultural context of Indigenous peoples then led to inclusion of the axiological and ontological dimensions. More recently, Chouinard and Cram (2020) looked at cul-

turally responsive practice across three program and community contexts (Western/North American, Indigenous, and international development). This led to the inclusion of a dialogical dimension based on the centrality of language in the construction of meaning, knowledge, and interpretation. There are currently 10 dimensions where culture and evaluation interconnect: epistemological, ecological, methodological, political, personal, relational, institutional, axiological, ontological, and dialogical.

These 10 dimensions are dynamic and overlapping aspects of culture that come together and merge throughout an evaluation, shaped by the program and community context, diversity of stakeholders engaged in the process, political agendas, the evaluator's cultural positioning (e.g., insider-outsider), strength of relationships and community connections, institutional agenda, and dedicated time and resources. As culture is a complex concept, each program context will have its own unique cultural ecology while, at the same time, being enmeshed in a hierarchy of social, political and cultural forces (Guzman, 2003) outside of the immediate program context. Thus, each evaluation context and program context is shaped by what Massey (2005) might refer to as "a multiplicity of trajectories" (p. 119) that are always in dynamic tension and transformative potential with their surroundings (both locally and more broadly). Most importantly, the evaluation context operates within a politically active and highly contested space (Lefebvre, 1991) that at its most fundamental level is shaped by the discipline of evaluation, itself a historically situated and highly institutionalized practice. While each of our 10 dimensions highlight a specific aspect of culture, as a totality, the dimensions are dynamic and continually interweaving throughout the evaluation process. In so doing, they highlight aspects of culture that have the potential to open up new possibilities for change; in other words, the dimensions form what

Soja (1989) might consider a “transformative dynamic” (p. 122).

Ontological

The ontological dimension looks at the nature of reality as a multifaceted construct deeply implicated in how we make meaning, and in whose meaning counts as valid in an evaluation. Cultural responsiveness requires an understanding that those with power in our society can create and legitimize their own reality while marginalizing others. Not all versions of reality are considered equal. This is the foundation on which evaluators are “helicoptered” into communities, where they are expected to work despite coming from one cultural world and being imposed on another. Culturally responsive evaluators should not hesitate to interrogate a program’s cultural fit and identify mismatches as risks for funders.

However, many people cannot see the boundedness of their own worldviews, their own realities; privilege is blind to itself. Culturally responsive practice thus first requires evaluators to critically assess their own worldviews and versions of reality to ensure that cultural context is not lost beneath the weight of their own version of reality. A key piece of this is an understanding that evaluation itself reflects a Western understanding of reality, and it may not “fit” with an Indigenous world that is relationships: with one another, with the environment, and with the cosmos (Wilson, 2008). The negotiation of the space between in order to be in relationship is a spiritual activity, often guided by cultural rituals of first encounter (Irwin, 1994).

The arguments of different Indigenous peoples based on spiritual relationships to the universe, to the landscape and to stones, rocks, insects and other things, seen and unseen, have been difficult arguments for Western systems of knowledge to deal with or accept. These arguments give a partial indication of

the different worldviews and alternative ways of coming to know, and of being, which still endure within the Indigenous world. (Smith, 2012, p. 74)

Epistemological

The epistemological dimension reflects the diversity of approaches to knowledge construction, with particular attention given to how knowledge is co-created between evaluators and stakeholders and among stakeholders themselves. Attention is also given to the evaluator’s role in navigating what are often a multiplicity of competing inquiry paradigms and beliefs about how knowledge is constructed and which/whose knowledge is of most value. For evaluators, this often means decisions need to be made about whose voices to include and which stakeholders to select. Many culturally responsive evaluators consider such decisions largely methodological and therefore requiring a technical response in the selection of the right method choice. At its extreme, competing methodologies might be balanced within this response in an effort to bridge funder and community priorities. Leal (2010) identifies this methodological fix as “technocracy” (p. 95)—that is, a process that reduces what are fundamentally epistemological and ontological issues to what is measurable and procedurally knowable, thus creating a process of depoliticization that ultimately preserves the status quo.

By contrast, the epistemological dimension raises critical questions about the tools, techniques, and instruments we use to address concerns of culture in our work as evaluators. It reframes the ways in which we construct knowledge with others and the kinds of knowledge we co-construct in evaluation contexts (what does it include–occlude). And it interrogates what influences, shapes, and demarcates our co-constructions. The challenge of the epistemological dimension is for culturally responsive evaluators to move

beyond technocracy to see the construction of knowledge as a cultural endeavor, in which language plays a key role as a conduit of worldviews (see “Dialogical” below). For Indigenous peoples, the construction of knowledge occurs within the context of relationships, and this is surely also a lesson culturally responsive evaluators explicitly or implicitly know wherever they are working.

Ecological

The ecological dimension highlights the temporal nature of context, drawing attention to evaluation as a process that is not fixed but unfolds over time, relationally connected to larger sociopolitical systems that act on and influence the local program and program setting historically, politically, culturally, and economically. There are multiple spatial and temporal dynamics at play in the evaluation process. As Massey (1993) has argued, “‘space’ is created out of the vast intricacies, the incredible complexities, of the interlocking and non-inter-locking, and the network of relations at every scale” (p. 161). Understanding the cultural context of the program and community shifts the evaluator’s gaze from the local to the global, from knowledge and understanding a community’s needs, priorities, and aspirations to a dynamic layering of the many interconnected influences that continue to shape the history, culture, sociology, and politics of the local context.

As evaluators, we walk into spaces already infused with their own cultural rhythm and dynamics that are constantly being shaped and reshaped by the people who live and interact and interconnect within them; there are no neutral spaces. Cultural responsiveness requires an understanding of a community’s history, culture, and background, especially as the community’s ecology will influence the evaluation design, process, and consequences (Ormond, Cram, & Carter, 2006). Building collaborative relationships and engaging community stake-

holders in the evaluation process can thus provide evaluators with an understanding of context, history, and cultural influences that are essential for culturally responsive practice, including the structural analysis of barriers to program success. Thus, one of the key components of the ecological dimension is cultural and contextual clarity, as the need to determine what is salient within a particular context, how culture is lived and experienced, provides evaluators with a more thorough understanding of what evaluation might look like locally.

Methodological

The methodological dimension describes the range of philosophical approaches to knowledge construction that guide evaluators in their work, and that influence the methods and instruments selected to collect data and analyze results. This dimension highlights the fact that all methodologies, no matter how collaborative or transformative they may seem, remain social, cultural, historical, and political constructions (Hopson, 2003). For evaluators working in culturally responsive program contexts, funder or client accountability-focused requirements often conflict with community input, needs, and concerns. While the tension between funders and evaluators may seem to be confined to small evaluation moments, the tension is nested within a broader discourse about whose worldview is privileged (and whose are lost or silenced) within these Western evaluation paradigms (see “Epistemological”).

In this regard, we are drawn to Reagan’s (1996) notion of “epistemological ethnocentrism” as a way of describing the contestability of knowledge that is mediated by social, political, and cultural influences, all of which are deeply embedded in the process of social inquiry. This dimension also has a strong cultural dynamic in terms of methodological justifications of validity (Kirkhart, 1995, 2005) and commensurability with the

local cultural community and program context. Evaluative data needs to be understood in terms of the community's cultural reality, bringing the focus on the use of collaborative approaches to methodological and method choice to enhance data quality and increase the validity and trustworthiness of evaluation findings. Cultural responsiveness also leads evaluators to push for inclusiveness through methodology. Genuine inclusiveness is more than an invitation to the table; it recognizes how an invitation and subsequent engagement operate both culturally and within a context of inequities in the distribution of societal power and privilege.

Political

The political dimension is focused on the relationship between evaluation and politics (Cronbach, 1980; House, 1993; Weiss, 1993), which in culturally diverse communities has methodological and relational expression, as questions about equity, access, self-determination, and social justice dominate. Questions about who participates and who does not, and whose values shape the evaluation dominate, as issues of power and privilege shape the boundaries and possibilities of practice in culturally diverse evaluation contexts. A key part of being culturally responsive means making explicit the interconnections between social, political, historical, and cultural influences within the evaluation context, and asking questions about the cultural relevance and commensurability of methods and methodologies that have been developed from outside of the community. It may also mean negotiating power imbalances, especially between communities and evaluation funders.

The use of collaborative approaches where evaluators and stakeholders work in partnership throughout the evaluation, though necessary in culturally responsive practice, cannot be assumed to bridge issues of power and privilege that accompany estab-

lished (and privileged) knowledge systems. The political dimension thus finds expression in relationships among stakeholders, evaluators, and funders within the context of the evaluation, in terms of the barriers and biases that may exclude people from participating in the evaluation process, and more broadly in terms of the norms and values that influence the legitimacy of certain approaches of knowledge construction over others. Collaborations that involve funders and decision makers, alongside communities, may also potentially bridge political divides that restrict the acceptability of evaluation findings that initiatives work for communities by virtue of strengthening their cultural capacity and supporting aspirations for equity, social justice, or decolonization. Central to this is the promotion by evaluators of "the inherent sovereignty and humanity of Indigenous peoples" (Kwaymullina, 2016).

Personal

The personal dimension positions the evaluator as an engaged accomplice in the evaluation process, who, in partnership with stakeholders, co-constructs evaluative knowledge. From a culturally responsive perspective, being rooted in the evaluation process requires a heightened "critical subjectivity" (Heron & Reason, 1997), not only of themselves and their own cultural standpoints but also of the epistemological and ontological assumptions and biases that guide their personal and professional practice. According to Rogoff (2003), "To learn from and about communities other than our own, we need to go beyond the ethnocentric assumptions from which we each begin. Often, the first and most difficult step is to recognize that our original views are generally a function of our own cultural experiences, rather than the only right or possible way" (p. 24).

This "multilateral self-awareness" (Symonette, 2004) requires self-understanding

within a particular context, and an understanding of self as evaluator/inquirer positioned within a dominant Western and privileged paradigm. Cultural responsiveness calls on evaluators to reflect on their cultural norms, personal beliefs, identity, biases, values and privileges, and to critically reflect on what frames their thinking, seeing, judging, and being as evaluators. As it is difficult for anyone to be an exact “fit” within a cultural ecology, so evaluators’ understanding of self enables them to negotiate the multiple levels in which they might be insiders and outsiders and to be explicit about where they stand when they offer their services to a community (Smith, 2012) (also see “Relational”). While this process of self-reflection and critical self-positioning is important, often it is a shared life experience and cultural familiarity with community stakeholders that is the foundation of culturally responsive practice.

Relational

Relationships define the culturally responsive evaluation context, as evaluators and stakeholders work together in partnership to define and work within the methodological parameters of the evaluation context. Stakeholders can include those who work closely with the program or program beneficiaries themselves, and depending on the evaluator’s familiarity with the community, can include a cultural guide to advise of protocols and ease an evaluator’s journey into the community. In all culturally responsive contexts, understanding and listening are considered key relational dimensions, as evaluators strive to move toward more trusting, respectful, and open relationships. Spending time both formally and informally (e.g., as a participant in community events) in the community building relationships is essential, as it provides a way to become familiar with community norms, values, and customs. In this way, culturally responsive evaluators signal their commitment to

the strengthening of authentic, trustworthy relationships.

A key characteristic of this dimension is evaluators’ identity or self-identity, as their positioning along the insider–outsider continuum frames their epistemology, their methodological decisions, and their approach to questions of ethics (Herr & Anderson, 2014). Challenging simple dichotomous assumptions is essential, especially given the heterogeneity and intracultural variability that exists within groups, further obfuscating insider–outsider binaries. Borrowing Patricia Hill Collin’s (2000) term *insider within* illustrates the complexity of evaluators’ dichotomous positionalities, as no matter which community they come from, they are also positioned as researchers or evaluators educated in Western modernist ideals, even though they might not enjoy all of the “rights” of that position either.

Institutional

The institutional dimension brings focus to the overarching professional norms, ideologies, and institutional policies and practices that guide evaluation practice, especially those focused on norms of accountability that exert incredible pressure and influence on culturally responsive practice. This dimension conveys a sense of what are often incompatible agendas and information needs between community members and funders, with evaluators seeking to balance conflicting priorities. Also included in this dimension are differences in evaluation purpose and rationale, with evaluators most often holding aspirations for cultural responsiveness that are related to equity, social justice, and decolonization, and seeking to adopt collaborative approaches to ensure they are working with the community rather than on it. This dimension also includes the time and resources necessary to build relationships, understand the program and its cultural context, and prepare for the evaluation, all of which can be

more challenging to manage in culturally complex settings where relational responsiveness may require an evaluator to be patiently present as a community moves at its own pace.

Axiological

The axiological dimension is focused on questions of ethics, with explicit emphasis on relationships between evaluators and program participants in an evaluative context. Cultural responsiveness requires that evaluators build relationships with stakeholders and with program and community participants as a way to understand the cultural context of the community and ensure the validity and integrity of the evaluation. This dimension also brings focus to the ethical implications of culturally responsive practice, as evaluators are often motivated by a sense of responsibility to ameliorate injustice and promote principles of fairness and equity, as well as by a strong belief in the inclusion of perspectives and voices from communities often left out of the conversation.

The use of collaborative approaches is considered essential in culturally responsive practice, as these have the potential to counterbalance external discourses of accountability, support empowerment, and build community capacity. From a cultur-

ally responsive perspective, understanding ethics in social inquiry means transforming it from its traditional position of neutrality to an aspirational meliorism inspired by values, moral principles, and a desire for social justice. For evaluators, these values and principles are not necessarily found in evaluation texts and guidelines as they themselves are culturally embedded (Smith, 2012), although some guidance is offered in the alignment of Weber-Pillwax's (1999) notions of respect, reciprocity, and responsibility, with characteristics of good practice highlighted by Chilisa (2020), Reviere (2001, in Chilisa, 2020), and Smith (2012) (Table 13.2). This demonstrates that being able to offer evaluation practice aligned with local values and cultural practices about how people engage with one another is fundamental to culturally responsive Indigenous evaluation.

Dialogical

The dialogical dimension acknowledges that language is central to understanding culture, bringing awareness to translation, interpretation, and representation in evaluation texts. When there are differences in language and communication styles between evaluators and stakeholders and community members, attention is also given to the role of translators in shaping mean-

TABLE 13.2. Indigenous Ethical Principles and Practices

Weber-Pillwax (1999)	Chilisa (2009)	Smith (2012)	Reviere (2001, in Chilisa, 2009, p. 412)
Respect	Deep respect for spirituality	Having a love for the people	Maintain inquiry rooted in strict interpretation of place
	Importance of agreement and consensus	Being a face that is known	
Reciprocity	Dialogue, particularity, individuality and historicity	Looking, listening . . . then speaking	Work to legitimate the centrality of African ideals and values
	Iterative process of many circles	Being humble	
Responsibility	Researcher as transformative healer	Sharing, hosting, and being generous	Uncover hidden, subtle, racist theories embedded in methodologies

ing, as they often play an active interpretive and ethnographically significant role in the evaluation and in the construction of knowledge. Translation is thus considered more than a technical process, as the process itself raises questions about authorship, authority, power, status, relationship, and voice—all of which are immensely relevant and powerful in culturally responsive program and community contexts. This dimension also includes differences in formal and informal structures and styles of communication and levels of literacy, bringing focus to adapting approaches to accommodate local communication styles and language literacy. Also acknowledged is the West's power to shape, frame, and define words (and processes) associated with the power to define and shape meaning, what Delpit (1988) refers to as a "culture of power" (p. 280). The imperative falls to culturally responsive evaluators to frame language in culturally appropriate ways with the community and that this make sense to the local community itself.

CONCLUDING REMARKS

In order to make sense of Indigenous people's lived experiences, we evaluators need to have an analysis of the disruption of Indigenous people's sovereignty and our ability to fulfill our caretaking roles and responsibilities for the land. Indigenous well-being, vitality, and sustainability cannot be ensured without the restoration of this relationship with the land, through the return of Indigenous territories (Tuck & Yang, 2012). Such an analysis leads to an understanding of the complexity of Indigenous lives, including the burden of intergenerational trauma and current disparities. This is the foundation of a strengths-based sociocultural analysis where structures are seen as needing to change, rather than the focus being on Indigenous people needing to assimilate or integrate into a culture that is not their own. As Kawakami and colleagues (2007, p. 323) write, "By decolonizing *evaluation* method-

ologies, we aim to re-center ourselves within our own lands."

In some colonial societies, it may be easy to ignore this plea for culturally responsive Indigenous evaluation because a discourse of Indigenous deficits and the need for Indigenous peoples to "change" to fit in is often unchallenged common sense. In other colonial societies, it may be easy to ignore this plea because Indigenous peoples are invisible or unrecognized as citizens. As Waziyatawin (Wilson) and Yellow Bird (2005, p. 1) note, "current institutions and systems are designed to maintain the privilege of the colonizer and the subjugation of the colonized, and to produce generations of people who will never question their position within this relationship." However, this needs to change, as time is short, both for Indigenous peoples and for the well-being of a planet that supports us all.

The journey through this 10-dimensional framework for culturally responsive evaluation practice may have left you a little breathless by the fullness of what it asks of you as an evaluator—both personally and professionally. Tying you down and asking you to be culturally responsive necessitates that you look first at yourself and ask, "Who am I?" as a way of beginning to unpack the multiple cultural ties and influences that inform the person you've become. This is not a one-off exercise, as we continually ask this question as our roles change, as we shift location and affiliations, and as we navigate ages and stages on our own life course. Most importantly though, having answers to this question will enable you to be more open in your offer of your evaluative services no matter where you choose to practice. The knowledge arising from culturally responsive Indigenous evaluation inquiry can support Indigenous community development, self-determination and nation building, and decolonization. As such, it is good evaluation practice that many communities—including non-Indigenous communities—may welcome (Cram, 2018). As Kwesi Prah (1999, in Chilisa, 2020, p. 415) (Ghana) has

written, “We cannot in all seriousness study ourselves through the eyes of other people’s assumptions.”

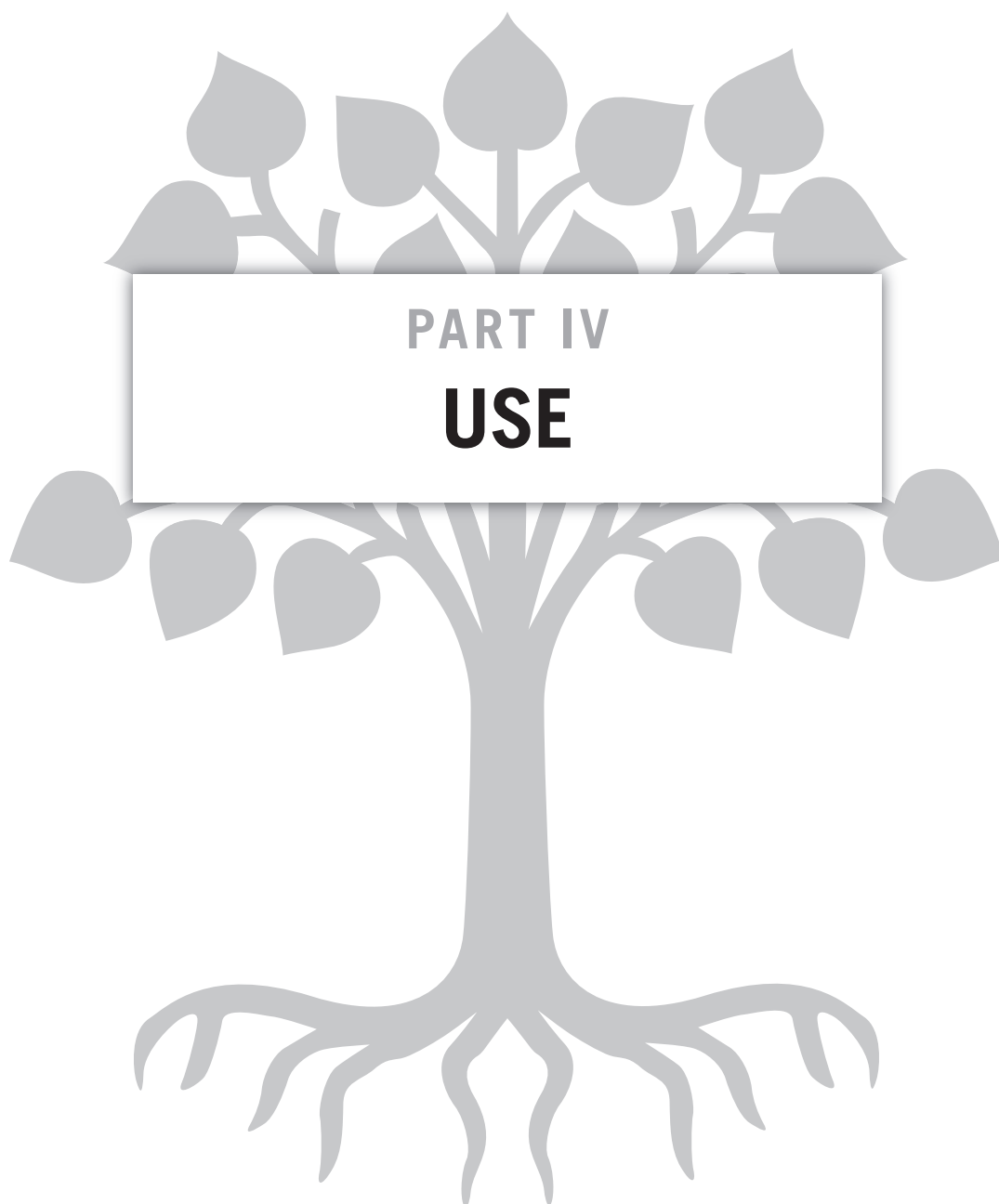
While we have said that lived experience is fundamental to culturally responsive evaluation practice, so is an authentic commitment to gaining a lived experience within the context of a cultural ecology. The good news is that you do not have to journey alone, as there are those both within the evaluation profession and out in communities who will be willing to team up and guide you. They will give you courage to flex and adapt the evaluation tools you know and to seek out new ones that are a better fit for collaborative methodologies. If you are new to the histories of Indigenous peoples who invite you to work with them, then finding out these histories may make you weep. This is when colleagues will give you hand up and get you back on your learning pathway. Too many tears have already been shed by those whose access to the goods and resources of this world have been restricted because they are not themselves the right fit. What they need is your skills and your commitment to representing them well within evaluation so that they can continue to advocate for wider responsiveness to their needs, priorities, and aspirations.

REFERENCES

- Anderson, W. (2007). The colonial medicine of settler states: Comparing histories of indigenous health. *Health and History*, 9(2), 144–154.
- Brant Castellano, M. (2004, January). Ethics of Aboriginal research. *Journal of Aboriginal Health*, pp. 98–114.
- Chilisa, B. (2020). *Indigenous research methodologies* (2nd ed.). SAGE.
- Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation*, 30(4), 457–494.
- Chouinard, J. A., & Cram, F. (2020). *Culturally responsive approaches to evaluation*. SAGE.
- Chouinard, J. A., & Hopson, R. (2016). Toward a more critical exploration of culture in international development evaluation [Special issue]. *Canadian Journal of Program Evaluation*, 30(3), 248–276.
- Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). Routledge.
- Cram, F. (2018). Conclusion: Lessons about Indigenous evaluation. *New Directions for Evaluation*, 159, 121–133.
- Cram, F., & Mertens, D. (2015). Transformative and indigenous frameworks for mixed and multi method research. In S. N. Hesse-Biber & B. B. Johnson (Eds.), *The Oxford handbook of mixed and multiple methods research* (pp. 99–109). Oxford University Press.
- Cram, F., & Mertens, D. (2016). Negotiating solidarity between indigenous and transformative paradigms in evaluation. *Evaluation Matters—He Take Tō Te Aromatawai*, 2, 161–189.
- Cronbach, L. J. (1980). *Toward reform of program evaluation: Aims, methods and institutional arrangements*. Jossey-Bass.
- Delpit, L. D. (1988). The silenced dialogue: Power and pedagogy in educating other people’s children. *Harvard Educational Review*, 53(3), 280–298.
- Frierson, H. T., Hood, S., & Hughes, G. B. (2002). Strategies that address culturally responsive evaluation. In J. Frechtling (Ed.), *The 2002 user-friendly handbook for project evaluation* (pp. 75–96). National Science Foundation.
- Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). A guide to conducting culturally responsive evaluations. In *The 2010 user friendly evaluation handbook* (pp. 75–96). National Science Foundation.
- Guzman, B. L. (2003). Examining the role of cultural competency in program evaluation: Visions for new millennium evaluators. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 167–181). Erlbaum.
- Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, 3(3), 274–295.
- Herr, K., & Anderson, G. (2014). *The action research dissertation: A guide for students and faculty* (2nd ed.). SAGE.
- Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. *New Directions for Evaluation*, 92, 31–43.
- Hood, S., & Hopson, R. K. (2008). Evaluation

- roots reconsidered: Asa Hilliard, a fallen hero in the “nobody knows my name” project, and African educational excellence. *Review of Educational Research*, 78(3), 410–426.
- Hopson, R. K. (2003). *Overview of multicultural and culturally competent program evaluation: Issues, challenges and opportunities*. The California Endowment.
- House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. SAGE.
- Irwin, K. (1994). Māori research methods and processes: An exploration. *SITES*, 28, 25–43.
- Kawakami, A., Aton, K., Cram, F., Lai, M., & Porima, L. (2007). Improving the practice of evaluation through indigenous values and methods: Decolonizing evaluation practice—returning the gaze from Hawai’i and Aotearoa. In P. Brandon & P. Smith (Eds.), *Fundamental issues in evaluation* (pp. 219–242). Guilford Press.
- Kirkhart, K. E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice*, 16(1), 1–12.
- Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In S. Hood, R. K. Hopson, & H. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 21–39). Information Age.
- Kwaymullina, A. (2016). Research, ethics and Indigenous peoples: An Australian Indigenous perspective on three threshold considerations for respectful engagement. *AlterNative: An International Journal of Indigenous Peoples*, 12(4), 437–449.
- LaFrance, J., & Nichols, R. (2009). *Indigenous evaluation framework: Telling our story in our place and time*. American Indian Higher Education Consortium.
- LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *Canadian Journal of Program Evaluation*, 23(2), 13–31.
- Leal, P. A. (2010). Participation: The ascendancy of a buzzword in the neo-liberal era. In A. Cornwall & D. Eade (Eds.), *Deconstructing development discourse: Buzzwords and fuzzwords* (pp. 89–100). Practical Action.
- Lefebvre, H. (1991). *The production of space*. Blackwell.
- MacDonald, B., & Parlett, M. (1973). Re-thinking evaluation: Notes from the Cambridge authority, teacher education and educational studies. *Cambridge Journal of Education*, 3(2), 74–83.
- Massey, D. (1993). Politics and space/time. In M. Keith & S. Pile (Eds.), *Place and the politics of identity* (pp. 139–159). Routledge.
- Massey, D. (2005). *For space*. SAGE.
- Mertens, D. M. (2009). *Transformative research and evaluation*. Guilford Press.
- Nobles, W. (1991). Extended self: Rethinking the so-called negro self-concept. In R. L. Jones (Ed.), *Black psychology* (3rd ed., pp. 295–304). Cobb & Henry.
- Ormond, A., Cram, F., & Carter, L. (2006). Researching our relations: Reflections on ethics and marginalisation. *AlterNative: An International Journal of Indigenous Scholarship*, 2(Special supplement: Marginalisation), 180–198.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). SAGE.
- Reagan, T. (1996). *Non-Western educational traditions: Alternative approaches to educational thought and practice*. Erlbaum.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.
- Sefa Dei, G. J. (2006). Introduction: Mapping the terrain—Towards a new politics of resistance. In G. J. Sefa Dei & A. Kempf (Eds.), *Anti-colonialism and education: The politics of resistance* (pp. 1–24). Sense.
- Senese, G. (2005). The PENAL project: Program evaluation and Native American liability. In S. Hood, R. K. Hopson, & H. T. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 129–147). Information Age.
- SenGupta, S., Hopson, R. K., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation*, 102, 5–19.
- Smith, L. T. (1992). Te Rapunga i te Ao Marama: The search for the World of Light. In *The issue of research and Māori*. Monograph No. 9. Research Unit for Māori Education, University of Auckland.
- Smith, L. T. (2012). *Decolonizing methodologies: Research and Indigenous peoples* (2nd ed.). Zed Books.
- Soja, E. (1989). *Postmodern geographies: The reassertion of space in critical social theory*. Verso.

- Stake, R. E. (1975). To evaluate an arts program. In R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 13–31). Merrill.
- Symonette, H. (2004). Walking pathways toward becoming a culturally competent evaluator: Boundaries, borderlands, and border crossings. *New Directions for Evaluation*, 102, 95–109.
- Tuck, E., & Yang, K. W. (2012). Decolonization is not a metaphor. *Decolonization: Indigeneity, Education and Society*, 1(1), 1–40.
- Waziyatawin, & Yellow Bird, M. (Eds.). (2005). *For Indigenous minds only: A decolonization handbook*. University of New Mexico Press.
- Weber-Pillwax, C. (1999). Indigenous research methodology: Exploratory discussion of an elusive subject. *Journal of Educational Thought*, 33(1), 31–46.
- Weiss, C. H. (1993). *Evaluation* (2nd ed.). Prentice-Hall.
- Wilson, S. (2008). *Research is ceremony. Indigenous research methods*. Fernwood.



PART IV

USE

CHAPTER 14

Practical Policy Evaluation

Aligning Theory with Practice

Eleanor Chelimsky

INTRODUCTION

Working as an evaluation practitioner for more than half a century now, I've found, over time, a number of important areas where I haven't been able to rely on existing evaluation theory to guide my steps. Sometimes I believe this occurred because theory, in a young field, had not yet explored those areas; other times, it was because the exploration had been inadequate or incomplete in various ways; still other times, it would seem that the contextual complexities involved in trying to keep desired ends in view while improving the evaluative means to achieve them were just too great (Chelimsky, 2013). In any case, I worked to develop my own approaches in these areas and also to clarify and treat their ramifications as they emerged. So, in response to Marv Alkin and Tina Christie's request that I try to summarize today what may be most useful in my own experience for other evaluation practitioners, I propose to review here some of my efforts to increase the ability of our evaluation processes to deal with varying

governmental and political environments, and then examine which among them may have ongoing relevance for current evaluation practice.

Looking back, I see three important places where contemporary evaluation theory failed me: (1) in its misunderstanding of the likelihood—indeed, the *normality*—of unexpected political interventions in the evaluation process; (2) in its assumption that evaluation findings would achieve routine use in policymaking; and (3) in its unawareness of the implications that a legislative environment might have for evaluation practice.

THREE PROBLEMS FOR EVALUATION PRACTICE: 1965–1980

Unexpected Political Interventions in the Evaluation Process

When I started doing evaluations at North Atlantic Treaty Organization (NATO) in the 1960s, and producing various required reports, I noticed that during the publica-

tion process, my work seemed to have been subject to deletions that altered the tenor of my analysis or conclusions. Since this happened a number of times, I began to worry about an evaluation process that allowed major changes to a final report without consultation with the author and without discussion of any such situation in the theoretical literature. With the noteworthy exception of Max Weber at the end of his life, most theorists were predicting an easy, uneventful integration of evaluation with policymaking.

This issue arose again on my return to the United States and an evaluation job at the MITRE Corporation, where I came face to face with those same theoretical views of how the use of evaluation findings would occur “naturally” among government’s program developers and policymakers. This, of course, conflicted with what I had already observed of external constraints on research at NATO, as well as similar problems I was encountering in my new work with different executive branch agencies in the United States. So I asked MITRE to sponsor a symposium to look at various actual *experiences* of policy or program integration of evaluation, and especially, the degree to which both evaluators and government executives felt that evaluator independence had been appropriately considered (Chelimsky, 1977). The discussions at the symposium clarified some of the political issues constraining use, and also confirmed a second issue.

Theoretical Assumptions about the Actual Use of Evaluation Findings

A premise of our theory in the 1970s, then, was that use would occur as a matter of course in government given strong methodological quality in the evaluation. But as Tom Cook (1995, p. 104) remarked at the American Evaluation Association’s Vancouver conference in 1995: “Twenty-five years ago, many evaluators naively anticipated that their results would be routinely used as the central input to policy decisions”

(cited in Chelimsky, 1997, p. 104). Indeed, at MITRE’s 1977 symposium, evaluation managers in government had pointed out that, in their experience, use had *also* depended on political vulnerabilities surrounding the kinds of question being asked and on the degree to which the evaluative answers actually met those managers’ needs. The discussions indicated a direct relationship between the purpose of the evaluation (i.e., the kind of question being asked), and the degree or type of use its findings received. Analyzing those discussions, I found three general user or evaluator perspectives that implied different types of use, and characterized these in the symposium report (Chelimsky, 1977, Vol. I) as follows:

- A knowledge perspective
- A management or development perspective
- An accountability perspective

In the case of the knowledge perspective, it seemed that the kind of use often experienced was indeed likely to have been related not only to the credibility and methodological quality of the study but also to whether the results accorded with what the agency user had originally been hoping to find. With respect to the management or development perspective, use seemed to have been tied to not only the closeness and appropriateness of the evaluator’s answers to the manager’s questions but also an open, trusting relation between evaluator and manager that allowed the manager’s experience to be fully integrated in the study. Finally, for accountability work to be used, managers again noted the importance of credibility in the study’s methodological quality, but just as importantly, of harmony (or at least lack of disharmony) in the findings with respect to what the managers were hoping to establish (e.g., “looking good” relative to public inquiries about the technical quality, timeliness, integrity, or objectivity of the agency work).

To me, this meant that, with respect to use, the *kind* of evaluation question involved took on a new importance. Management and development types of evaluations were clearly more likely to be used, while accountability and knowledge evaluations could very well remain on the shelf if they uncovered the “wrong” information. As such, this second issue with theory—the need to think of different evaluation perspectives both in responding to different kinds of user questions and in expectations about use—deepened my original concern about the true political constraints on the evaluator, since it confirmed the relationship of a particular political context to the policy use of findings.

I recognized a third theoretical gap when I started developing an evaluation unit—the Program Evaluation and Methodology Division, or PEMD—at the U.S. Congress’s Government Accountability Office (the GAO) in 1980.

Implications of a Legislative Context for Evaluation Practice

Evaluation theory, I discovered, had concerned itself less with the legislative environment for evaluation than with the agency environment I had been accustomed. This, of course, meant a new learning process for me, to understand what appeared to be different concerns about the structure of democratic government, different (legislative) pressures and functions, a different (daily) presence of competing public interest values, and different manifestations of corporate and lobbyist power. More importantly, it meant the need to strategize about the changes these differences might involve for a new evaluation unit working in this environment.

Government structure, for example, was a much more frequent subject of congressional concern than what I had seen reflected in agency discussions. In particular, this concern involved two of Montesquieu’s pre-Enlightenment ideas:

1. How the historic failures of democratic institutions were related to the rise of “factions” (special interests that worked against the public interest, corrupted democratic processes, and increased social tensions)
2. How the separation of powers, along with checks, balances, and the congressional function of oversight, presented the best way to avoid such failures

Himmelfarb (2008) reminds us of how important these ideas had been to America’s founding fathers, and also notes that Jefferson filled 28 pages of his *Common-place Book* with extracts from Montesquieu, “more space than he devoted to any other author” (p. 199). But I found to my surprise that these venerable ideas were heavily present in conversations I had on the Hill (*in 1980*) about why the evaluation of public programs and policies (and especially, the accountability of executive agencies) had now become so important to the U.S. Congress.

My first thoughts here were that I would need to think about increased evaluative emphasis in four legislative areas that seemed to be impacted by these structural considerations: (1) accountability evaluation to support congressional oversight; (2) analysis of statements made during confirmation processes in which candidates were grilled for special-interest involvement; (3) examination of obstacles to the congressional function of consensus building and compromise that allow for the regular legislative management of political conflict (Jewell & Patterson, 1966); and (4) credibility defenses for our evaluative work to thwart possible interventions by Montesquieu’s “factions” or special interests.

Even more alive, however, than this worry about democratic decline was the warring nature of the *everyday* legislative environment, based on alternative, legitimate, but often conflicting public interest values. This was, of course, very different from the single-strand Democratic or Republican val-

ues engagement (depending on the most recent election) that typically underlay policymaking in executive branch agencies and which had imparted at least a perception of tranquility to executive branch evaluations (Chelimsky, 1998). For me, the constancy of these legislative branch confrontations—and their ferocity—meant that our evaluations, throughout the duration of their entire process, would need to consider, prepare for, and be able to answer likely objections, based on legitimate but conflicting public interest values. I expected those objections also to emanate from the three types of constantly existing pressure I'd already seen affecting project or program evaluations for executive agencies:

- Pressure from cross-branch politics (this derives from the checks-and-balances architecture of the federal government, and I had often observed its effects on accountability evaluations of agency programs)
- Pressure from *bureaucratic* politics (this typically had to do with management relationships within the user agency and arguments about strategies or tactics)
- Pressure from the dominant *professional* culture of the sponsor agency—whether it be law, science, audit, engineering, or something else—and its ability to deal with the concepts and recommendations of evaluation (Chelimsky, 2009)

It turned out that the legislative context would add considerably to cross-branch pressures because of a general executive agency resentment of congressional oversight, which would often result in hostility, protective secrecy, silos, and sometimes irrational divisions and separations of data collection for the same subject area across agencies.

Adding further to this already turbulent political climate was the dynamism of legislative thinking and policymaking brought by the institution of congressional hearings

(and the public debates and values battles that informed them). These added a potential for sudden change that I hadn't seen in the relative stability of agency policy development, and it meant that important disruptions could happen unexpectedly not only to proposed new programs and policies, or those up for reauthorization, but also, of course, to their evaluations. Yet evaluation theory continued to tell us that the nature of program change is "slow and incremental" (Shadish, Cook, & Leviton, 1991, p. 188); that no single authority can radically change a program (Shadish et al., 1991, p. 39); and that "timeliness is an overrated concern" (Cronbach et al., 1980, p. 63).

Finally, I noticed (1) a more *overt* presence of lobbyists compared to what I had seen at NATO or in work for American executive branch agencies and (2) a much closer connection to public opinion, especially for House members, brought by primaries and elections. This sent me back to Montesquieu (since special interests threaten not only the institutions of democracy but also evaluation findings and their use), and the need to establish strong credibility defenses not only for our findings, but for our entire evaluation process.

All of this deeply entrenched political context appeared to promise a changed role for evaluation—not only expanded and more powerful but also more defensive—compared with what I had thus far experienced. I imagined this role might require, first, an evaluation practice that paid much closer attention to political balance, transparency, and the need to eliminate any appearance of advocacy in an environment of legitimately conflicting public interest values; and second, a much more comprehensive inclusion, for each evaluation, of the study of the relevant past history—in particular, (1) the social, economic, anthropological, psychological, military, cultural, and/or political trends and events that had brought us to the current specific legislative position dictating the evaluation question; (2) the

probable ramifications of this history for our evaluation thinking, especially for the kinds of evidence needing to be produced, and the likelihood of use; and (3) revelations about the types and strength of special interest pressures likely to be directed at our work and the specific credibility defenses we would need. What I didn't realize at the time was that this kind of legislatively generated expansion might not signify merely another set of problems to resolve, but that it could, in the long run, have a positive effect, moving evaluation beyond the uncertain support it usually received within the agency environment to a more foundational importance for government generally.

So, what in fact turned out to be important for us evaluation practitioners, working within these theoretical gaps in a new legislative environment?

CHANGES TO EVALUATION PRACTICE IN PEMD, 1980–1994

Looking back at the nearly 300 evaluations we produced for the Congress between 1980 and 1994, it's easy to see that they all, in one way or another, included efforts to adapt an existing evaluation process to a new political context. Some of these efforts worked quite well, others helped but only over time (after, say, a new election), and still others courted disaster because of risk taking but found policy use anyway; some ended up waiting on the proverbial shelf, others failed lamentably. On average, however, there was remarkable use of our findings, and it seems to me that there are four changes in practice, among the many we implemented over time, that helped greatly in achieving that use:

1. Recognition of an ongoing need to develop systematic credibility defenses against political or methodological attack
2. Increase in evaluator knowledge of the relevant history and its relation to the policy context of the current proposal

3. Attention to complex spillover issues embedded in evaluation questions
4. Much more detailed analysis of differential user perspectives in an enlarged stakeholder cohort

DEVELOPING CREDIBILITY DEFENSES

We discovered early on in PEMD that we would need three kinds of credibility defenses to protect against various types of political attacks on our work and that the effort to develop and implement these defenses was never-ending because the political context and protagonists were continually changing. The three kinds were (1) reputational strength, (2) substantive/methodological quality, and (3) presentational effectiveness.

1. For reputational strength, I looked to:

- A stellar advisory board, presenting both research and subject-matter expertise
- Publications by staff members in various social scientific journals
- Careful attention to the staff balance of Republicans and Democrats
- Networking with social science departments in various universities
- The presence of staff from the relevant social science branches, possessing methodological capabilities, an ability to synthesize as well as analyze, and a bent toward innovation and risk taking
- National and international efforts to contact and learn from the experience of other evaluation units working in contexts similar to our own
- Just having a *reputation* for excellence, it turned out, was enough to deter at least some political attacks

2. We found that substantive/methodological quality in a credibility context involved persuasiveness in an evalua-

tion report with respect to what has been learned, how it has been learned, and what caveats need to be imposed on that learning. As a result, establishing this kind of quality requires methodological expertise, subject-matter expertise, and also objectivity and nonadvocacy in the evaluators. One problem we had to deal with here was tension existing between substantive and presentational credibility, between telling all, in the necessary methodological detail, and being read. We treated this difficulty (a) by using technical appendices (which support substantive credibility and can be as long and tedious as necessary) and (b) by supplementing reports with oral briefings. These briefings were usually quite well accepted, avoided the burden of reading time for legislators, and also brought us stronger findings through the willingness of policymakers to share their own experiences with us.

3. Presentational credibility's focus is, of course, generally on readability but it also involves concern for reasonable length, precision of logic and style, and the avoidance of technical language in the report (to the degree possible). President Carter's "plain English" rule applied here, certainly, but I also dreamed of Pascal's test of eloquence: that is, saying or writing things in such a way that our readers would not only understand and be convinced by them, but become so caught up in them that they actually did something about them. As time passed, we learned that neutrality in the language we used—for example, no blaming and shaming about past failed legislation—could be decisive with respect to the eventual reception and use of our findings.

These three aspects of credibility are, of course, tightly connected: The reader's knowledge of the evaluators' reputation for expertise, along with the demonstration of that expertise in what is reported, and the lack of pomposity in the way it's described and presented, together affect that reader's perception of the evaluation's objectivity

and competence. We found that they helped greatly in our work, both by deterring a lot of the special interest attacks I worried about and by promoting the eventual use of our findings in this new climate of overtly competing public interest values.

LOOKING AT PAST HISTORY

I had learned at NATO that merely looking at the precipitating factors generating a new study was never enough: Longer-term trends leading up to the relevant policy issues were always heavily involved and important to understand in detail. In our work at PEMD, I developed two kinds of review that we would perform based on the character of the evaluation. The first was the typical review of the evaluation literature, whose purpose was to help us avoid perpetuating past mistakes in design, and which included interviews with the earlier evaluators to find out what the obstacles to their performance had been, where they would have done something different today, and how political or other arguments had manifested themselves. We did this review for most evaluations. But then we added another kind of three-sided political review, based on my NATO experience, which I called the Forces Framework, which we used only for complex, politically hot questions with a long historic buildup (e.g., I did this for our initial look at environmental problems; see U.S. Government Accounting Office (USGAO), *The Nation's Water*, PEMD-86-6). This kind of review looked beyond evaluator experience to the study's historical context, in terms of three driving forces:

1. A precipitating event (or events)
2. The current political balance with respect to public interest values (this balance was typically cyclical, involving, say, an ongoing values conflict such as the one opposing Hamiltonian big spending to Jeffersonian parsimony)
3. Longer-term issues that are not cyclical, but always underlying to some degree

(e.g., racial or gender inequality, or gradual changes in public attitudes)

Today, I would use this three-way review to look at police reform, for example. Precipitating events are the denial of individual rights by police; the current political balance chiefly twists between the civil rights of Black people and community “law and order”; and finally, longer-term issues (involving slavery, institutional racism, the development of camera technology, and the eternally complex realities of police traditions and functions): All are important factors for establishing evaluative understanding and for developing a credible public interest strategy.

This three-way study of the past, done at the beginning of a politically volatile evaluation, helped us in a number of ways. First, it provided the basis for understanding what the political arguments about the proposed study were likely to be, and hence to build that understanding into evaluator thinking right from the start. Second, it helped identify all the stakeholders implicitly or explicitly connected to the later use (or non-use) of our findings. Third, it allowed better targeting of the evaluation question, based on overall past experience, with respect to whether a proposed study might *feasibly* accommodate a knowledge, accountability or management/development perspective. Fourth, it allowed a quick response from us as to the likely usefulness of the study (here we would focus on issues like the following: Is the problem diagnosis persuasive, given past history? Are all the relevant voices included? Have important theoretical or other unresolved evaluative disagreements been recognized or is there overemphasis of one view versus another? Is there evidence of special interest involvement or of the “single narrative” embedded in the proposal (Chelimsky, 2012)? More recently, I used it to help forecast problems of public engagement (e.g., a likely resistance to proposed dietary improvements in a public health effort to combat obesity) that called for more sophisticated policy-action strategies,

based on past success or failure (Chelimsky, 2019).

We learned that some problems that appeared to be matters of poor communication in a new policy or proposal could in fact originate in much deeper issues involving opposed but legitimate public interest values. As Isaiah Berlin (cited in Bailyn, 2015b, p. 235) once pointed out, it is often democratically appropriate:

to coerce men in the name of some goal (let us say justice, or public health) which they would, if they were more enlightened, themselves pursue, but that once one claims to know what others need better than they know it themselves, one is in a position to ignore the actual wishes of men or societies, to bully, oppress, or torture them in the name and on behalf of their “real” selves . . . albeit often submerged and inarticulate.

Resistance to policy, especially when it is scientifically driven policy, may seem frivolous, but it is often based on real fear, can often be predicted, and always needs to be taken seriously in policy-action strategies.

Over time, we found that, as evaluators armed with this kind of research, we were better able to resist political pressures, survive the turbulent legislative climate, design stronger studies, and improve the likelihood of their use. Not always, of course; we were ignominiously defeated, for example, by the National Rifle Association and the medical device industry. Also, we noticed that the presence of this research at the *beginning* of an evaluation brought about two transformations in our work process: (1) Negotiation with sponsors, which had in the past occurred only at the end of a study, now instead began at the start (or even before) and continued throughout the evaluation, and (2) we became much more tentative in our suggestions for modification; that is, we saw them less as applications of principle than as efforts to achieve balance in this environment of continuously conflicting public interest values where we knew we had a lot to learn.

I should mention that the Forces Framework was an attempted response to

Moynihan's point that the failures of the community-action programs of the 1960s were much less due to their "intrinsic qualities" than to unmeasured political factors exogenous to the programs: that is, to a "rise of upper middle-class white disaffection with the direction of American society, occurring in conjunction with an even more powerful surge of the civil rights movement, associated with an inevitable and long overdue rise in militant black assertiveness" (1969, p. 202). It was a way for us to better understand those "unmeasured" political factors and try to account for them in timely fashion.

COMPLEX NEW QUESTIONS AND THE USE OF THE CHECKLIST

One of the first things I had noticed in discussing 1980s legislative thinking about evaluation on the Hill was the importance attached to prospective questions. Since these had not been part of the retrospective fare I was used to in my agency work, we had to focus immediately on new methods (like the prospective evaluation synthesis or the interrupted time series) to help us answer questions about the likely future of new legislative proposals or to make estimates based on the state of knowledge existing in an emerging subject area (Chelimsky, 1987). But as time went on, we began to get more and more questions involving extremely complex subject areas (like the safety net or defense) where we recognized major untreated issues of spillover. To help us answer these new types of questions, we turned to the checklist and used it as a sort of Ariadne's thread with which to confront the Minotaur and his mazes, in general, and subjects like climate change or program/policy sustainability, in particular (Chelimsky, 2019).

These spillover checklists brought three evaluation practice issues to the fore: (1) the need to document—at the evaluation design stage—expected interactions between the

subject area and other related systems or fields of knowledge, with special emphasis on any conflicts in policies across two agencies or areas; (2) the need to determine objectively the likely relative importance of those interactions to the evaluation; and (3) the need to identify any existing data sets, relating to any of those interactions, that might be available for use. In addition, they sometimes showed the importance of including new areas of expertise in our analyses, a requirement I dealt with initially either by bringing in outside specialists or borrowing experts from the relevant agencies.

For some evaluations, this analysis was quite limited because whatever spillover existed had only trivial importance for the study. But for others, it was significant. For example, this kind of analysis showed the possibility of merging databases from both health care and law enforcement agencies to better understand the incidence of accidental shootings nationally, and again, in a study of high school dropouts, the utility of combining education and police data to clarify the size of the problem. Soon we saw that answering questions about carbon emissions in the United States always requires the involvement of at least three agencies (Environmental Protection Agency [EPA], Energy, and Treasury) and also requires multiple kinds of expertise—scientific, economic, financial and regulatory, for example—in the evaluation team, along with the usual methodological skills. And, of course, in areas of public interest value conflicts, when proposed policies battle across tribal borders—as in, for example, the problem of convincing culturally different social groups to change their behavior to avoid disease—we had to learn how to develop a difficult synthesis of voices that could include both the powerful and the destitute (USGAO, AIDS Education: Reaching Populations at Higher Risk, PEMD-88-35).

With respect to the checklist on sustainability issues, I targeted, in particular, the need for evaluators to oppose, modify, or at least negotiate, formulations that:

- Ignore cyclical conflicts in public-interest values.
- Involve specific provocations to public habits and tastes.
- Fail to consider regional experience and trust issues based on that experience.
- Cross multiple bureaucratic boundaries.
- Compete unnecessarily with other natural or human systems.
- Neglect relevant practitioner experience.
- Disrupt strong community traditions (Chelimsky, 2019).

From the viewpoint of practice, we found that using checklists in this way (1) helped us deal with the increasing complexity of some questions; (2) reminded us of important, sometimes half-forgotten earlier evaluative work that brought new clarity to hidden agendas; and (3) allowed us to negotiate successfully some appropriate changes to proposed legislation. From a theoretical viewpoint, however, the checklists often served to identify a new and “verifiable world of interconnections, of relationships, which together added up to a different and better picture of the whole” (Bailyn, 2015a, p. 95).

EXAMINING USER PERSPECTIVES IN AN EXPANDED COHORT

As I mentioned earlier, the legislative context added a new set of hostile users to the typical evaluation user paradigm when the question involved executive agency accountability and congressional oversight. This problem is basic. My experience was that Executive branch agencies tend to reject congressional oversight, which they see as an intrusion into their world of work by unknowledgeable outsiders. For us, this meant that, in such cases, we often had findings that were targeted for implementation by Executive branch agencies, despite the fact that these agencies could clearly be expected to oppose them. This difficulty

turned out to be so profound, and so much a part of the government structure—which *encourages* agency independence as a part of separation of powers—that, when our findings were in fact used, it was often because of an atypical achievement of trust and understanding among the managers and staff involved. What we did learn was not to make the problem worse, to pay careful attention to agency sensitivities and long-term experience and knowledge, and to court their willingness to experiment with reasonable modifications. I developed a strategy for promoting use that involved eight stages:

1. Negotiation with all users in planning a use-focused evaluation
2. Evaluation design to accommodate user information needs and time frame
3. Agreement on final evaluation question(s), timing, and use
4. Performance of the evaluation
5. Informal, separate interim briefings to congressional sponsors and agency users
6. Review of the evaluation report by all prospective users
7. Monitoring (by us) of reception and use of the evaluation’s findings
8. *Live* postmortem with users to assess lessons learned. (Chelimsky, 1983, p. 161)

Over time, we learned the value of supporting our findings and modification requests with a great deal of evidence. But even in hostile situations, we found that this expanded analysis of user thinking and experience brought us both increased understanding and more intelligent use of our findings.

SOME THOUGHTS FOR TODAY’S PRACTITIONERS

As I think about how the changes I’ve outlined here in PEMD’s evaluation practice of the 1980s and early 1990s correspond to today’s environment for evaluation (especially the speed and size of Internet impacts, the normalized presence of disinformation,

the increased political pressures to produce the “right” evaluation findings, the loss—or trivialization—of compromise as a basic mechanism for legislative function, and the sometimes staggering, globalized complexity of the new questions now being posed to evaluation), it seems that it might perhaps be helpful to focus on three important issues.

Maintaining an Awareness of Whole-Problem Complexity

First, as everyone knows, it’s usually not only necessary to break large societal problems down into manageable segments in order to analyze and evaluate them, but also to keep a compensating vision of the entirety of the problem going forward. For example, when considering whether, say, an employment and training program deserves high expectations for decreasing poverty, it’s important not to get stuck in buzzword types of contrasts, but to keep in mind something like William Julius Wilson’s description of the many factors involved in poverty: “discrimination, economic deprivation, poverty concentration, isolation from traditional role models, inadequate education, collapsing labor market opportunities, the emergence of an underground economy” (Wilson, 1994, p. 48, cited in Chelimsky, 1996). Policymakers and evaluators are often surprisingly naive about what a new proposal—Just Say No, for example, or No Child Left Behind—can be expected to accomplish; a careful synthesis of what is known about all relevant aspects of the problem can be immensely helpful.

Continue Expanding Strategies to Promote Use

Second, the use of findings in *any* democratic climate is never easy. Today’s environment not only requires considerable evaluator courage, modesty, and persistence but also a new stress on use-promoting efforts, such as frequent regular meetings with even remote users; more accurate—that is, less absurdly optimistic—expectations about

timing and goals; willingness to research past evaluative failures and their causes; demonstrable evaluator objectivity with regard to subject matter; careful integration of user knowledge and experience; and a much better understanding of how to target policy-action strategies for behavior change. Recent public health failures in seeking to change dietary habits, for example, or to get people to wear masks to protect against COVID, remind us again that conflicted public interest values are difficult to treat, and that we are still utterly inept at achieving acceptance by the public of unpopular scientific or evaluative findings. This is not a new problem—the example of Prohibition and its continued failures with respect to alcoholism in the society comes immediately to mind—and it’s not a simple one. We need to expand our understanding of cultural synthesis, improve the ambient public trust in science, make more intelligent choices of policy-action strategies, and take more interest in the art of messaging.

Staying Flexible about the Role of Evaluation

Third, my guess is that if policymaking eventually returns to traditional modes of development across Executive, Legislative, and Judicial branches of government, evaluation should continue to be useful to policymakers in its usual roles, that is, “providing information for discrete decisions, acting as a means for managing conflict and promoting social change, stimulating program staff to critically examine their assumptions and behavior, or acting as a societal ritual whose function is to calm the citizenry and perpetuate an image of government rationality” (Floden & Weiner, 1978). But I think an important addition to those roles, especially if the normalization of disinformation should continue, comes from Jacques Ellul’s (1964) prescient view of what we evaluators must not sacrifice or forget as we work in today’s technological society: “As long as the first stage of analysis is incomplete, as long as the problems are not correctly stated, it

is useless to proffer solutions. And, before we can pose the problems correctly, we must have an exact description of the phenomena involved" (p. xxxii). To do that, of course, evaluators will need to face both inward and outward: inward to consolidate our field and outward to draw more closely on the knowledge of other fields. Kuttner (2020, p. 74) has put it very well:

Reading articles by writers from a variety of scholarly traditions, I'm often struck by how leading economists or sociologists or political scientists, working on similar problems from different disciplines, are unfamiliar with one another's work. This is a casualty of the fragmentation of the social sciences, and it impoverishes discourse. Albert Hirschman, educated as an economist, spent the latter part of his career as a moral philosopher. He read widely across disciplines, puckishly calling himself a "trespasser." If we are to make sense of the great questions of our age where politics meets economics, we need more trespassers.

As trespassers, we need to sharpen and better document our vision of how the global world works and how evaluation fits into that world; and we need to marshal the strength and creativity to apply that vision to new problems of sometimes global dimensions.

Even so, however, given the complexity of what lies in front of us, we are likely to have to face quite a few failures, to live with sometimes very small gains, and to adapt to the mobility of a political universe that not only constrains and burdens but also legitimizes, our work. A way to help with that is to keep alive the memory of evaluation's successes and failures of the past, the sense of one step building on another, toward the necessary cumulation of knowledge and Ellul's "exact description of phenomena." Over the longer term, political environments are volatile, power changes hands—appropriately and legitimately—in a democracy, and evaluation needs to be defensible across the spectrum of political mixes, and

especially as the mix changes. Despite all that, I expect evaluation to remain a critically important tool for policymakers, and a source of joy and freedom for evaluators, even if changing the world with one study is not an option. As Camus (1955) said, "The struggle itself toward the heights is enough to fill our hearts. We must imagine Sisyphus happy."

REFERENCES

- Bailyn, B. (2015a). History and the creative imagination. In *Sometimes an art: Nine essays on history*. Knopf (Borzoi).
- Bailyn, B. (2015b). The search for perfection. In *Sometimes an art: Nine essays on history*. Knopf (Borzoi).
- Camus, A. (1955). *The myth of Sisyphus and other essays* (J. O'Brien, Trans.). Knopf.
- Chelimsky, E. (Ed.). (1977). *Proceedings of a Symposium on the Use of Evaluation by Federal Agencies* (Vols. I & II). MITRE Corporation.
- Chelimsky, E. (1983). Improving the cost-effectiveness of evaluation. In M. C. Alkin & L. C. Solman (Eds.), *The costs of evaluation*. SAGE.
- Chelimsky, E. (1987). Retrospective and prospective analysis: Linking program evaluation and forecasting. *Evaluation Review*, 11(3), 355–370.
- Chelimsky, E. (1996, Summer). *From incrementalism to ideology and back: Can producers of policy information adjust to the full spectrum of policy climates?* Center for Urban Affairs and Policy Research, Northwestern University.
- Chelimsky, E. (1997). Thoughts for a new evaluation society. *Evaluation*, 3(1), 97–118.
- Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation*, 19(1), 42–44.
- Chelimsky, E. (2009). Integrating evaluation units into the political environment of government: The role of evaluation policy. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice*. *New Directions for Evaluation*, 123, 51–66.
- Chelimsky, E. (2012). Valuing, evaluation methods, and the politicization of the evaluation process. In G. Julnes (Ed.), *Promoting valuation in the public interest: Informing policies for judging value in evaluation*. *New Directions for Evaluation*, 133, 77–83.

- Chelimsky, E. (2013). Balancing evaluation theory and practice in the real world. *American Journal of Evaluation*, 34(1), 91–98.
- Chelimsky, E. (2019). Improving the match between sustainability questions and evaluation practice: Some reflections and a checklist. In G. Julnes (Ed.), *Evaluating sustainability: Evaluative support for managing processes in the public interest*. *New Directions for Evaluation*, 162, 75–79.
- Cronbach, L., Ambron, S., Dornbusch, S., Hess, R., Hornik, R., Phillips, D., . . . Weiner, S. S. (1980). *Toward reform of program evaluation*. Jossey-Bass.
- Ellul, J. (1964). Author's foreword to the revised American edition. In *The technological society*. Knopf, Random House and Vintage Books.
- Floden, R., & Weiner, S. (1978). Rationality to ritual: The multiple roles of evaluation in governmental processes. *Policy Sciences*, 9(1), 9–18.
- Himmelfarb, G. (2008). *The roads to modernity*. Vintage.
- Jewell, M. E., & Patterson, S. C. (1966). *The Legislative process in the United States*. Random House.
- Kuttner, R. (2020, September 24). Can We Fix Capitalism? *New York Review of Books*.
- Moynihan, D. P. (1969). *Maximum feasible misunderstanding: Community action in the War on Poverty*. Free Press, Macmillan.
- Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation: Theories of practice*. SAGE.
- U.S. Government Accountability Office. (1986). *The nation's water* (PEMD-86-6). Author.
- U.S. Government Accountability Office. (1988). *AIDS education: Reaching populations at higher risk* (PEMD 88-35). Author.

CHAPTER 15

Context-Sensitive Evaluation

An Approach to Research-Based Use

Marvin C. Alkin

INTRODUCTION

I have spent over 50 years thinking about evaluation. Many of those early years, I didn't know very much and was not especially productive. But, early on when I was Director of the UCLA Center for the Study of Evaluation, I produced a so-called "CSE" (context-sensitive evaluation) model that focused on defining the various evaluation stages. This was first published in the long defunct journal *Evaluation Comment* and subsequently reprinted in a Carol Weiss book (Alkin, 1972).

The original CSE model (or prescriptive theory) defined the various stages of evaluation and prescribed what actions the evaluator should take in each stage. What was unique at the time was that it identified process and outcome phases at both the formative and summative stages. The emphasis on considering outcomes in formative evaluation may have been an early indicator of what I now call "summary formative evaluation" (Alkin & Vo, 2018, p. 3). Furthermore, the identification of process

measures at the summative stage, which was called "documentation evaluation," attested to the necessity of verifying and reporting exactly what activities constituted the program as it was operationalized.

I have seen prescriptive theories that, when written, were allowed by their authors to continue in the same form for many years, while theory writers constantly benefit from new writings and experiences that influence their thinking. When Ernest House (1977) was a visiting scholar at UCLA, I saw how he developed new insights into how he viewed evaluation based in large part on reading the works of two Belgian philosophers (Perelman & Olbrechts-Tyteca, 1969). Further, in a conference session I chaired, Egon Guba commented on a discussion by a student on his theoretical writings. He said, "You see, in many ways I am not the 'real' Egon Guba . . . that was the one writing seven to ten years ago" (Guba, 1979, p. 139), demonstrating that theories sometimes change over time in the minds of those who write them.

I got bored with the idea of developing and modifying prescriptive theories; I did

not want to write a stagnant theory. So, for many years, instead, I turned my attention to research on evaluation use, to the examination of comparative evaluation theory and to active participation in conducting evaluations.

CONTEXT-SENSITIVE EVALUATION

A further resurrection of my prescriptive theory world came about when writing the first and second editions of my textbook, *Evaluation Essentials*. I was forced to think about the small details of what I prescribed should be done when conducting an evaluation. By writing the text in 2009, it really forced me to “lay it out” (Alkin, 2011; Alkin & Vo, 2018).

My years of experience doing research on evaluation use made me more attentive to the specific factors affecting use. I categorized these within four contexts: the evaluator context, the stakeholder/user context, the organizational and social context, and the evaluation context. Thus, CSE does not refer to a simple context. CSE addresses four contexts.

In addition, more recently, I discovered Robin Miller’s 2010 article, “Developing Standards for Empirical Examinations of Evaluation Theory,” in which she maintained that evaluation theories are not specific enough, and she presented criteria for judging evaluation theories by the criterion of having achieved the evaluation goal in practice.

Based on Miller’s criteria, I sought to define criteria to better depict my theory. For me, this means (1) describing the range of primary applicability (program size and type of evaluation), (2) precisely defining what I mean by use, (3) showing the prescribed *activity* sequence of the evaluation that details the major activities engaged in when conducting the evaluation (further details on these are the evaluation prescription), then (4) designating research-based observable actions that lead to greater use.

In combination, these would provide users with a stronger basis for selecting the theory and for implementing it properly. It would also allow an observer of an evaluation presumably conducted using a CSE approach to verify to a certain degree that the theory had been implemented appropriately.

CONDITIONS OF PRIMARY APPLICABILITY

And so, let me start the further delineation of CSE. The CSE approach is most applicable to small- and medium-scale programs (primarily local and not federal or large-scale state programs). These types of programs require the kind of close and ongoing stakeholder relationships that I consider a necessary part of CSE, as it focuses on formative and summary formative evaluations.

The Goal

The *goal* of CSE is use. But this goal must be precisely defined. The high-quality information gathered must be responsive to and reflective of the needs of the stakeholder/user community. Aside from that, use can take on many forms. I further define *evaluation use* as making decisions, establishing or altering attitudes, or building an individual or organization’s evaluation capacity. Use might involve an interested primary user or other local users and lead to use through information acquired as findings and/or as a consequence of participating in the evaluation. Furthermore, the evaluation information must serve as a dominant (or, one of multiple) influences related to the program being evaluated in its entirety, its components, the role of individuals in any way associated with the organization, and the views and understandings of individuals in any way associated with the organization. Most importantly, I conceive of evaluation use as oriented toward *this program at this time*. Thus, I exclude the impact on “external users,” that is, those actions external

to the program, its funders, or its constituencies. The evaluator must be concerned about best serving the stakeholders and programs that are the recipients of his or her services. I specifically exclude the concept of “influence” of findings since it implies use at another time or in other programs. If the evaluation findings have influence elsewhere or are of use to other programs, fine, but that is an evaluation research finding and not the goal of CSE.

Activity Sequence

Now, what are the activities that constitute CSE? The activities would include many characteristics that are present in other theories. (Or, at least, I feel so strongly about these aspects of conducting an evaluation that I want to believe they are prevalent in other theories.) The activity sequence that I prescribe is mostly linear, but there are many overlapping activities:

1. Perhaps the most important aspect of conducting an evaluation is the presence of a competent, fully engaged, *use-dedicated evaluator*, a topic on which I have specific perspective and further discuss later in the chapter.
2. The evaluation must seek to gain full *understanding of who the stakeholders are*, both those primary stakeholders who are most likely to be in a position to influence the direction of the program being evaluated and those others whose voices should be heard.
3. The evaluator needs to *gain respect, trust, and credibility*. These are partially a by-product of developing positive stakeholder relationships.
4. The evaluator must gain an understanding of the *organizational/social and political context*.
5. Although not always possible, a description of the program and its activities leading to the *development of a logic model* is essential.
6. *Questions or issues* to be addressed should flow from an understanding of the logic model and be directed by *the specific concerns of stakeholders* since their buying into the questions is absolutely essential. The evaluator must thoroughly examine the questions to be sure that they are able to be used in the evaluation; that is, questions must meet appropriate technical, ethical, and political feasibility standards, and must be subject to potential modification as the program progresses.
7. The evaluator has a role in the development of instruments and to ensure that they appropriately respond to the questions. However, instruments should be discussed with stakeholders to ensure that the *measures are accurately depicting the questions* that they had helped develop and approve.
8. The evaluator and stakeholders need to develop *an evaluation plan for the collection and analysis of both process and outcome data* along with a timeline for the completion of these tasks.
9. The evaluator has particular management responsibilities for ensuring that the *evaluation proceeds in an efficient and ethical fashion*.
10. Data acquired through the evaluation process *must be valued*—that is, the meaning of the data must be determined with respect to and associated with the *questions being asked*.
11. Evaluation outcome results and process understandings need to be reported and *communicated to appropriate audiences* in a form that is *clear and understandable*.
12. Throughout this whole process, the evaluator must be careful to *adhere to the appropriate evaluation standards* and codes of behavior.

These then are some of the characteristics that constitute the CSE Activity Sequence and that I optimistically believe appear in most (or many) evaluation theories.

RESEARCH-BASED OBSERVABLE ACTIONS

So, what are the unique actions that distinguish CSE, and where do they come from? Some of these so-called “unique” factors are a fine-tuning of commonly occurring evaluator actions. Some are truly unique. Where do these unique factors come from? Like others, I have been very active in conducting research on evaluation use. That research is quite extensive. Studies I’ve conducted were focused on examining the particular factors that tend to be associated with the evaluation and its results being used. A summary of the derivation of factors from this literature is found in Alkin and Vo (2018, pp. 314–316). A fine-tuning of research evidence leads me to focus on a cohesive framework that comprises four main categories or contexts—The *Evaluator Context*, the *User Context*, the *Individual/Organizational/Cultural Context*, and the *Evaluation Context*.

The specific evaluation use research factors (Column 1, Table 15.1) within the four evaluation contexts are drawn from five major research compilations of evaluation use factors (see footnote to Table 15.1). The first of these is the evaluator context. These are factors directly related to the evaluator’s professional identity and his or her competencies, beliefs, and value positions on various evaluation-related issues. By far, the most important element in this context area is the evaluator’s personal commitment to use. This refers to the way the evaluator, in his or her actions and interactions, demonstrates a deep commitment to attaining use. Moreover, it must be a priority of such importance that the evaluator seeks to infuse it into users’ thinking. Another factor, within the evaluator context, is technical competence. The evaluators’ personal credibility is an essential part of the evaluation being viewed as credible and thus worthy of use. It is important to note that the credibility of the evaluator, while partially attained at the outset, is mostly acquired through the evaluator’s actions throughout the evaluation. Indeed, it is the values and attitudes

of the evaluator that frame the evaluation. And the way the evaluator interacts with stakeholders and involves them in the variety of evaluation activities and is politically sensitive helps ensure use. Another element within the evaluator context I have called *context awareness*. The evaluator must be aware of the culture and his or her own positionality in relationship to that context.

The user context comprises actions that the evaluator must take in relationship to actual and potential users. The evaluator’s personal commitment to use is reflected in his or her attempt to gain stakeholders’ attention to this priority. This starts with gaining a fuller understanding of stakeholders’ prior experiences with evaluation and how they might shape their current views. Another area within the evaluator user category is *influence capability*. This refers to the potential, current, or enhanced role that users might have in making decisions about program change or their ability to modify attitudes about the program. The evaluator must work to develop in stakeholders a positive attitude toward evaluation and a commitment to use. This involves, at minimum, a willingness on the part of evaluators to actively involve users in the conduct of the evaluation.

In the third category, individual/organizational/community context, we note that programs exist within their larger organizations and within communities. Each of these contexts and their constituent parts, rules, and expectations exert influence on an evaluation. Discovery of these factors is not enough. Attention to these factors by the evaluator as reflected in his or her actions is essential. In particular, the evaluator seeks to understand the various individual and organizational values.

The fourth category, the evaluation context describes several important attributes of the evaluation. In the evaluation, both procedures and outcomes must be viewed as credible. The communication within both the process and the various types of reporting must be of high quality. Finally, the eval-

TABLE 15.1. Evaluation Use Research Factors, Evaluator Actions, and Sample Observable Actions

Research factors	Prescriptions	Sample observable actions
A) Evaluator Context • Evaluator personal commitment to use • Evaluator technical competence • Evaluator credibility • Context awareness	A1) Evaluator personally and intensely commits to attaining evaluation use.	A1) a) The evaluator comments on and emphasizes his or her commitment to the attainment of evaluation use in meetings with stakeholders. The evaluator can be seen asking questions about potential use during the evaluation. b) The evaluator helps stakeholders to understand use possibilities and may be seen teaching or facilitating stakeholders to engage in an analysis of use alternatives.
	A2) Evaluator assesses his or her own technical competencies and determines areas in need of further development.	A2) a) The evaluator assesses actions required in the conduct of the evaluation and the requisite technical skills. The evaluator assesses his or her own technical capabilities and either develops missing requisite skills or provides appropriate technical assistance.
	A3) Evaluator engages in actions that build greater credibility.	A3) a) The evaluator develops credibility by showing integrity, doing what he or she says he or she will do, and communicating clearly and transparently about the evaluation's progress. b) The evaluator engages stakeholders as thought partners by offering a menu of possible routes to pursue in an evaluation while facilitating their consideration of the advantages and disadvantages of each. For example, the evaluator could state, "Well, let's think about that possibility for a moment. What has to be true for that to work? What are possible challenges? How can we address them?"
	A4) Evaluator works towards developing further understanding of the individual and organizational contexts and assesses his or her own values in relationship to the various contexts.	A4) a) The evaluator interviews all primary user stakeholders to better understand their values and snowball interviews various other relevant stakeholders. The evaluator seeks information that supports or disconfirms alignment between the various values of the stakeholders and the organization. b) The evaluator systematically examines his or her own personal values and convictions before beginning an evaluation so that he or she is aware of how these personal views might influence his or her observations and questions. Furthermore, he or she examines differences between his or her own perspectives and values and those of stakeholders. c) While the evaluator has a primary obligation for improving the local program rather than addressing broader societal social justice issues, the evaluator has an obligation to point out to stakeholders' social justice concerns <i>within</i> the program and what is morally right. This should be communicated with stakeholders throughout the evaluation.

(continued)

TABLE 15.1. *(continued)*

Research factors	Prescriptions	Sample observable actions
B) User Context - Initial attitude toward evaluation - User influence capability - Meaningful Participation - User commitment to use	B1) Evaluator seeks to understand prior evaluation experiences of stakeholders and how this affects their current attitudes toward evaluation.	B1) a) The evaluator seeks to become more aware of past program evaluations and understand stakeholders' prior experiences with evaluation through initial discussions. Questions such as the following tend to be helpful: "Were you involved in the process when the program was last evaluated? What was that like?" "What do you hope will be different about your experience this time?"
	B2) Evaluator identifies those who have or might have influence capability and helps stakeholders to understand the role they might play in decision making.	B2) a) The evaluator interviews program participants and community members to understand who benefits from a positive or negative evaluation of the program and how. He or she asks various stakeholders, "Who are the people who have a stake in the outcomes of the evaluation?" and "How are they already positioned to affect changes in this program?" b) The evaluator seeks to work with individual stakeholders/users who will (or might) have the ability to influence decisions within the program. The evaluator asks questions such as "What does 'evaluation use' mean to your program? What role might you have in influencing program change of any type during the evaluation or at the summary stages of evaluation? How might your role be enhanced?" c) The evaluator seeks out those who are less influential to hear their views and show how they might have a greater voice in program improvement. This can be done through requesting interviews and input from stakeholders who may not be directly involved in decision making but are impacted by the program.
	B3) Evaluator actively involves users in all major phases of the evaluation.	B3) a) The evaluator involves stakeholders in playing the major role in the development of the logic model, helping to develop questions for the evaluation, examining methods to be used, participating in all data collection activities, and reviewing evaluation findings.
	B4) Evaluator works on developing a positive attitude toward evaluation and a commitment to evaluation use in stakeholders.	B4) a) The evaluator seeks to engage stakeholders in priority-setting activities. These "reality testing" opportunities offer stakeholders and evaluators opportunities to discuss the relevance of evaluation questions and data collection measures. Questions that are used to structure these discussions include the following: "Why is this question important?" "Might the answer to that question provide support for program changes or for modified attitudes?" "Would you be satisfied that this measure (i.e., an exam, a survey, and interview) provides relevant evidence for use related to the question that you have asked?" b) The evaluator's approach is particularly defined by a priori valuing by stakeholders. The evaluator asks: "If I told you at the end, or some other summary point of the evaluation, that the level of accomplishment for this particular question was X, would you conclude that was an indication of success? What would you consider to be an appropriate level/standard for the program to be deemed successful?" c) When working with stakeholders in developing the program's logic model, the evaluator inserts a use component following each output and each outcome in order to sensitize stakeholders to instances where use might occur.

(continued)

TABLE 15.1. *(continued)*

Research factors	Prescriptions	Sample observable actions
C) Individual/Organizational/Cultural Contexts • Organizational and individual perspectives and values • Interorganizational elements ○ Interest in evaluation and perspectives of those at a higher organizational level ○ Relationship of the program to the broader or related programs in the organization • Organizational and programmatic arrangements ○ Development stage of the program ○ Type of evaluation • Preexisting evaluation policy and legal bounds ○ Written evaluation requirements ○ Other contractual obligations ○ Organizational evaluation policy ○ Fiscal constraints	C1) Evaluator seeks to understand the perspectives and values of individual stakeholders and the views generally held within the program.	C1) a) The evaluator interviews individuals at various levels in the organization to better discern their perspectives and understand the importance of recognizing differences in the various perspectives. b) The evaluator seeks to interview and gain understanding of the perspectives of a sampling of stakeholder groups that do not have voice in program decision making. The evaluator acts as an intermediary in voicing these views. c) The evaluator seeks to understand organizational values and mores. This understanding is obtained through the interview process and should be communicated to stakeholders to ensure mutual understanding.
	C2) Evaluator seeks to understand the interests and values of those at other organizational levels and uses that information to enhance the evaluation.	C2) a) The evaluator seeks to understand how the program fits with the broader organization's mission. What is its relationship to other programs within the broader organization? What makes it unique or similar to other programs that organization offered? The evaluator might determine this by interviewing those in the broader organization and in other related programs within the organization. b) The evaluator asks questions such as "Who does the head of this program report to?" • "What is your understanding of how he/she perceives the program?" • "Have you had any indication of concerns that he or she or other organization administrators feel about the program?"
	C3) Evaluator accommodates the evaluation to the stage of the program and the type of evaluation.	C3) a) The evaluator develops understanding of the program's history and development by asking questions such as • "How did the program come about?" "What needs were intended to be addressed?" • "How have the program's goals and objectives evolved from what the program initially intended to where it is now, if at all, and why?" b) The evaluator focuses evaluation questions on and uses study designs appropriate for the program's stage. Doing so requires the evaluator to understand the program's scale (e.g., "Has the program been piloted? What was learned from that experience? At how many sites is the program currently offered? How has that changed over time? How many participants have completed, are currently engaged in, or have prematurely left the program?"). These questions should be asked in initial conversations with stakeholders and answered prior to designing the evaluation to ensure evaluation applicability.
	C4) Evaluator seeks to understand evaluation policy and legal bounds of the evaluation and ensures that they are adhered to.	C4) a) The evaluator checks and double-checks all evaluation requirements and contractual obligations to be certain that they are being met. The evaluator seeks to operate within these evaluation bounds. b) The evaluator seeks to understand organizational policy that might impact the evaluation. For example, the evaluator might ask, "Who is the contact person for arranging site observations?" Or the evaluator might provide a list of the various intended evaluation procedures and ask about approvals and organizational liaisons related to each.

(continued)

TABLE 15.1. *(continued)*

Research factors	Prescriptions	Sample observable actions
D) Evaluation Context D1) Evaluation Credibility D2) Communication Quality D3) Evaluation Usability	D1) The evaluation has high technical quality as appropriate to the situation. The evaluation organization has a very positive reputation. The evaluation has high relevance in that it meets the information needs of stakeholders.	D1) a) At the outset of the evaluation, the evaluator discusses other similar evaluations that the evaluation organization or the evaluator personally has successfully conducted. b) The evaluator seeks to select a design and technical characteristics appropriate to the evaluation context. c) The evaluator involves stakeholder in all aspects of the evaluation to ensure the potential relevance of evaluation information.
	D2) The evaluation communication, both informal and formal (oral and written) is clear, understandable, and trustworthy.	D2) a) The evaluator is constantly attentive to stakeholder communication preferences and recognizes the necessity of clarity in all communication relevant to stakeholders. The evaluator should establish preferred means of communication at the outset of the evaluation and reevaluate these means throughout to ensure effective communication. b) In providing summary evaluation reports, the evaluator includes a description of the program as it was actually implemented.
	D3) The conduct of the evaluation leads to the possibility of a high level of evaluation use.	D3) a) The evaluator takes a major role in ensuring the usability of the evaluation: personal dedication to use, understanding stakeholders' perspectives, engaging stakeholders in active participation in the evaluation, finding ways to teach stakeholders about use, and providing guidance in how to use evaluation findings.

Note. Research factors slightly modified from Alkin and Vo (2018, pp. 314–316). Based on research compilations by Leviton and Hughes (1981), Alkin (1985), Cousins and Leithwood (1986), Shulha and Cousins (1997), and Johnson et al. (2009).

uation must be a source of both conceptual and instrumental use.

THE PATH TO OBSERVABLE ACTIONS

The path from evaluation use research to observable evaluator actions is laid out in Table 15.1. In Column 1 is a list of research factors as derived from the five compilations of research findings cited. In Column 2 is the intermediate step of portraying each evaluation finding as an appropriate evaluator action. In Column 3 is the broad evaluation action statements as *specific* observable actions. These are actions that one might see or hear the evaluator engaging in when conducting a CSE.

I have depicted many examples of observable actions. Some of them, while constituting actions that I would want to observe in qualifying an evaluation as CSE, might, in fact, not be totally unique. Let me now comment on some actions that I believe are both extremely important aspects of CSE and particularly unique. I view these as essential elements of CSE.

The first major research factor is the *evaluator context*. In fact, the evaluator is a part of all four contexts. It is the evaluator's actions that complete the evaluation and lead to its potential use. This first major research factor relates to the evaluator as a person and his or her capabilities, commitments, and personal values. One of the most important observable actions in this context is the *dedi-*

cated personal commitment of the evaluator to evaluation use. I call this the *evaluator personal factor*. This can be shown by the way the evaluator continually stresses the importance of evaluation use as the mechanism for program improvement. The evaluator works with stakeholders to attempt to infuse in them the same dedication to use that he or she has (see A1.a. in Table 15.1). This recognition of the importance of stakeholders understanding use possibilities is accomplished by the evaluator seeking opportunities for teaching about evaluation use (See A1.b. in Table 15.1).

Two other evaluator actions within the evaluator context are especially important. The first deals with the evaluator's personal values and convictions. It is essential that the evaluator deeply consider their own values and how they might influence observations and interactions with stakeholders. This analysis must also include a search to understand differences between his or her own perspectives and values and those of the various stakeholders/users (see A4.b. in Table 15.1). Secondly, I have pointed out that CSE is focused primarily on small-scale formative and summary formative evaluations. Thus, the evaluation does not, and cannot, focus on solving broader social justice issues. The second is that the CSE evaluator's focus must be on improving *this* program at *this* time. However, the evaluator has an obligation to be aware of and point out to stakeholders any social justice concerns within the program and to point to directions that are morally and ethically right (see A4.c. in Table 15.1).

The second main research factor is the *user context*. Appropriate actions must be taken so that stakeholders/potential users recognize and act on their critical role in ensuring that use takes place. One essential action on the part of the evaluator is identifying stakeholders who might be in a position to use the information produced by the evaluation or by the actions engaged in by participating in the evaluation. The evaluator's role is not only identification but also

nurturing of stakeholders so they might have enhanced opportunities for evaluation use (see B2.b. in Table 15.1). While there are primary stakeholders whose opportunities for use are clear and well defined, stakeholders who are less influential and often ignored should have their voices heard. This is an obligation of the evaluator (see B2.c. in Table 15.1).

One aspect of the intent to gain meaningful participation is by asking stakeholders/potential users to reflect on the questions and measures of the evaluation inquiry. The examination of the importance of questions, the appropriateness of the measures and their relevance for potential use is an important component for gaining a commitment to evaluation use (see B4.a. in Table 15.1).

The evaluator in CSE involves users in all phases of the evaluation. An especially unique observable activity within CSE is that the evaluator engages stakeholders in a priori valuing when possible; that is, the evaluator works with stakeholders/users in assigning levels of outcome attainment or written descriptions of outcomes that would be considered "successful." This alerts stakeholders to the urgency and appropriateness of use and kinds of use that might be appropriate (see B4.b. in Table 15.1).

I previously noted that the development of logic models is a fairly common practice in evaluation. Stakeholders must come to recognize that the logic model is not only a means to better understand their program but also a vehicle for envisioning potential future avenues for change. In CSE, I propose that evaluators add a use component in the construction of the logic model after each output and outcome that may serve as a basis for considering future potential use (see B4.c. in Table 15.1).

The context-sensitive evaluator has an obligation to attempt to understand the multiple dimensions of the *individual/organizational/cultural context*. One aspect of this is reflected in the evaluator interviewing a variety of individuals to gain detailed understanding of their perspectives and

values (see C1.a. in Table 15.1). But understanding organizational perspectives and values of primary stakeholders within the organization is insufficient. The evaluator must seek out stakeholders who typically do not have a voice in decision making and act as an intermediary in voicing these views within decision-making contexts (see C1.b. in Table 15.1). Furthermore, the evaluator must seek to understand how the program fits within the broader organization and its mission (see C2.a. in Table 15.1).

There are several particularly relevant aspects to how the evaluation is conducted. One that derives from my very own early work focuses on a summary evaluation report. This report must include not only a description of the proposed program but also a description of how the program was actually implemented (see D2.b. in Table 15.1).

A final observable action with the category of evaluation research factors highlights the role of the evaluator as personally responsible for use. Personal dedication to use is reflected in the evaluation by finding ways to teach stakeholders about use and providing guidance in how to use evaluation findings (see D3.a. in Table 15.1).

In this chapter, I have sought to present a description of CSE that included a precise definition of the use goal, the circumstances of greatest applicability, a description of the four contexts of CSE, and some unique observable actions that would make the application of CSE recognizable in practice.

REFERENCES

- Alkin, M. (1972). Evaluation theory development. In C. Weiss (Ed.), *Evaluating action programs: Readings in social action education*. Allyn & Bacon. (Reprinted from *Evaluation Comment*, 1969).
- Alkin, M. (1985). *A guide for evaluation decision-makers*. SAGE.
- Alkin, M. C. (2011). *Evaluation essentials: From A to Z*. Guilford Press.
- Alkin, M. C., & Vo, A. (2018). *Evaluation essentials: From A to Z* (2nd ed.). Guilford Press.
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331–364.
- Guba, E. (1979). “The new rhetoric” of Egon Guba as compared to that of 10 years ago. *Studies in Educational Evaluation*, 5, 139–140.
- House, E. (1977). *The logic of evaluative argument*. Center for the Study of Evaluation.
- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
- Leviton, L. C., & Hughes, E. (1981). Research on the utilization of evaluations: A Review and synthesis. *Evaluation Review*, 5(4), 525–548.
- Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation*, 31(3), 390–399.
- Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research and practice since 1986. *American Journal of Evaluation*, 18, 195–208.

CHAPTER 16

The Ongoing Evolution of Utilization-Focused Evaluation Practice Informed by Theory, Research, and Reflection

Michael Quinn Patton

INTRODUCTION

Utilization-focused evaluation (U-FE) aims to support effective action and informed decision making based on meaningful evidence, thoughtful interpretation, and engaged deliberation. U-FE is a comprehensive decision framework for designing and implementing an evaluation to fit a particular situation and, in that situation, meeting the information needs of primary intended users to enhance their intended use of the evaluation. *U-FE is done for and with specific primary intended users for specific, intended uses.* This approach can be used for any evaluation in any context.

U-FE begins with the premise that evaluations should be judged by their utility and actual use. Evaluation's professional standards, guiding principles, and endorsed competencies all call for making evaluations useful. U-FE enhances use by facilitating the

evaluation process and designing any evaluation with careful consideration for how everything that is done, from beginning to end, will affect use. U-FE provides systematic, research-based guidance and a set of principles for deciding what approach to evaluation is most appropriate for a particular situation and specific primary intended users. U-FE is pragmatic and eclectic, so the U-FE toolkit encompasses every evaluation option methodologically, conceptually, theoretically, analytically, and processwise. Evaluation theorists, methodologists, and practitioners have generated an extensive, even daunting, menu of options to meet particular evaluation needs and demands, any of which can be made utilization-focused. U-FE doesn't prescribe what particular evaluation methods or approach to adopt but rather prescribes a process for determining how to conduct any evaluation with *unwavering attention to intended uses by intended users.*

In any evaluation, there are many potential stakeholders and an array of possible uses. U-FE requires moving from the general and abstract, from possible audiences and potential uses, to the real and specific: actual primary intended users and their explicit commitments to concrete, specific uses. The evaluator facilitates judgment and decision making by primary intended users. U-FE is personal and situational. The evaluation facilitator develops a working relationship with intended users to help them determine what kind of evaluation they need and will use.

A psychology of use undergirds and informs U-FE: Intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they've been actively involved; and by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way.

This chapter reviews the roots, evolution, and adaptation of U-FE over the last 45 years since the original research on use begun in 1975 that led to publication of the first edition of the book in 1978. Over the ensuing decades, U-FE has evolved, as has the field of evaluation in adapting to changes in the larger world. I begin with the current status of U-FE, then review the development of the approach as a foundational framework to guide practice.

THE STATUS AND INFLUENCE OF U-FE

Independent scholarly reviews of research on evaluation consistently find that U-FE provides the most fully developed theory explaining how to enhance use (King & Alkin, 2019) and is rated as most influential by evaluation practitioners (Becho, 2019; Haugh & Grodzicki, 2016). U-FE is the basis for the evaluation approach taken by

major philanthropic foundations (Christie & Lemire, 2019). A substantial body of research supports the premises and principles of utilization-focused evaluation (Alkin & King, 2016, 2017; Cousins, 2020; Ramirez & Brodhead, 2019).

U-FEs have been implemented around the world, from grassroots evaluations in Burkina Faso, West Africa (D'Ostie-Racinea, Dagenais, & Ridde, 2021), to international nongovernmental organizations (NGOs) such as the Global Alliance for the Future of Food (2020). Context matters. U-FE is highly sensitive to and must be adapted to context. That said, U-FE can be applied in any context where use is a priority. U-FE is not appropriate when a program is just going through the motions of evaluation because it is required and thus is undertaken with a compliance mentality. Where evaluation is viewed as merely a mandated paperwork exercise, as the pain that comes with receiving money from funders, U-FE becomes a burden because it demands thoughtful engagement. But when those involved in delivering, making decisions about, and funding programs want to learn, improve, and increase effectiveness, U-FE offers an energizing and results-oriented path forward. In a major review and synthesis of evaluation models and hypotheses on the nature of use, Contandriopoulos and Brousselle (2012) described U-FE as "the utilization paradise" (p. 70), the place where use flourishes.

EVOLUTION OF U-FE

The Beginning: Discovering and Conceptualizing the Personal Factor

The first edition of *Utilization-Focused Evaluation* was published in 1978 and featured research on factors affecting use based on case studies of 20 federal health programs in the United States. With colleagues and students in the University of Minnesota Evaluation Methodology Program, we interviewed the evaluators and those for whom the eval-

uations were conducted. We asked respondents to comment on how, if at all, factors extracted from the literature on diffusion of innovations and evaluation utilization had affected use of their study. Finally, we asked the respondents to pick out the single factor that each one felt had the greatest effect on how this study was used. From this long list of questions, only two factors emerged as consistently important in explaining utilization: (1) political considerations and (2) a factor we called “the personal factor.”

This latter factor was unexpected, and its clear importance to our respondents had, we believed, substantial implications for the use of program evaluation. The personal factor is the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates. Where such a person or group was present, evaluations were used; where the personal factor was absent, there was a correspondingly marked absence of evaluation impact. Use is not simply determined by some configuration of abstract organizational dynamics; it is determined in large part by real, live, caring human beings. Once understood, this became the foundation of U-FE. Thus, the challenge of increasing use comprises two parts: (1) finding and involving those who are, by inclination, information users, and (2) training and incentivizing those not so inclined.

Defining Use: Intended Use by Intended Users

The second edition of *Utilization-Focused Evaluation*, published in 1986, offered, for the first time, a definition of utilization: *intended use by intended users*. This specified the desired outcome of a U-FE.

Building on discovery and conceptualization of the personal factor, primary intended users are people who have a direct, identifiable stake in the evaluation. Identifying them at the start of an evaluation (and continuing to work with them as an evaluation progresses) is critical to ensuring that an evaluation is utilization-focused. Put

simply, without the engagement of primary intended users, there is no U-FE.

Many decisions must be made in any evaluation. The purpose of the evaluation must be determined. Concrete evaluative criteria for judging program success has to be established. Methods have to be selected and time lines agreed on. All these are important issues in any evaluation. The question is, “Who will make these decisions?” The utilization-focused answer is primary intended users of the evaluation. Clearly and explicitly identifying people who can benefit from an evaluation is so important that evaluators have adopted a special term for potential evaluation users: *stakeholders*.

The practice implication of focusing on intended use by intended users is that the early work of an evaluation process is to identify primary intended users out of the larger pool potential stakeholders, then organize primary intended users in a way that the evaluator can engage with them in exploring a menu of potential evaluation uses. This requires facilitating consideration of options of possible primary intended uses of the evaluation for these primary intended users. Evaluation questions, design, data collection, interpretation, and reporting all follow from intended use by intended users.

Discovering Process Use

The third edition of *Utilization-Focused Evaluation*, published in 1997, introduced *process use*. When I established the Minnesota Center for Social Research in the mid-1970s, I began the practice of following up every evaluation we conducted to find out how it was used. Those evaluations are the basis for many of the case studies and stories in my writings. Part of my preparation for doing each new edition of *Utilization-Focused Evaluation* is reviewing client feedback from evaluations and workshops. When, in the mid-1990s, I went to prepare the third edition of the book and began reflecting on what had happened in the field in the 10 years since the previous edition, I was struck

by something that my own myopia had not allowed me to see before. When I have followed up my own evaluations over the years, I have inquired from intended users about actual use. What I would typically hear was something like this: “Yes, the findings were helpful in this way and that, and here’s what we did with them.” If there had been recommendations, I would ask what subsequent actions, if any, followed. But beyond the focus on findings and recommendations, what they almost inevitably added was something to the effect that “it wasn’t really the findings that were so important in the end; it was going through the process.” In reflecting on that feedback, I came to realize that *the entire field had narrowly defined use as use of findings*. We have thus not had ways to conceptualize or talk about what happens to people and organizations as a result of being involved in an evaluation process: what I have come to call “process use” (Patton, 1997, 1998, 2007).

This idea of process use draws our attention to individual changes in thinking and behavior among those involved in evaluation as a result of the learning that occurs during the evaluation process. Changes in program or organizational procedures and culture may also be manifestations of process use. Process use includes helping those we work with learn to think evaluatively. Evaluative thinking is often alien to nonevaluators. Evaluative thinking includes the value we place on clarity, specificity, and intentionality; being systematic and making assumptions explicit; operationalizing program goals; distinguishing inputs and processes from outcomes; valuing empirical evidence; and separating statements of fact from interpretations and judgments. When we take people through a process of evaluation, at least in any kind of stakeholder involvement or participatory process, they often learn how to think in these ways.

The theory-to-practice implication is that, in working with primary intended users, process use expands the menu of options available from just different purposes for

using findings to a variety of options for engaging intended users in the processes of U-FE.

Globalization of Utilization-Focused Evaluation

The fourth edition of *Utilization-Focused Evaluation*, published in 2008, gave significantly increased attention to international and cross-cultural factors that affect use. As evaluation has increasingly become a global profession and enterprise, cultural competence has emerged as an issue affecting the credibility of evaluators to engage in diverse contexts with diverse stakeholders. This concern deepened the importance of beginning U-FE with a situation analysis identifying the contextual, cultural, organizational, political, social, and economic factors that affect how evaluation findings are used for program improvement and decision making.

Utilization in Complex Dynamic Systems

Essentials of Utilization-Focused Evaluation, published in 2015, brought systems thinking and complexity theory into U-FE. In understanding context to inform situation analysis when working with intended users, the *Essentials* book distinguished simple, complicated, and complex situations. In working with intended users to consider these distinctions, it is illuminating to engage them in discussion about what aspects of what they do are relatively simple, relatively complicated, and relatively complex.

The challenge I took up in the *Essentials* book was how to portray U-FE as a series of sequential steps, while also capturing the complex nature of the utilization-focused process as nonlinear, interactive, dynamic, and adaptive. Presenting U-FE as a series of steps was driven by the necessarily linear and sequential nature of writing and, to some extent, our human thinking processes involved in figuring out how to do something. But the utilization engagement process is not neatly linear and sequential.

There are interactions among the steps, feedback loops, recursive iterations, interrelationships, and interdependencies among those involved in the evaluation, and other complex dynamics that can be observed in any emergent system. For example, while identifying primary intended users occurs at the beginning of a U-FE, that process may be ongoing (Do we have the right intended users as evaluation questions emerge?) and new intended users have to be identified when there is turnover among those originally involved. In essence, the utilization-focused evaluator is asked to operate within two perspectives at the same time, one simple, stepwise, and neatly linear, the other complexly interactive and dynamic.

Utilization Heuristic

How do we manage complexity? Decision sciences have been identifying decision heuristics that cut through the messy, confusing, overwhelming chaos of the real world, so that we can avoid analysis paralysis and take action. We rely on routine “heuristics”—rules of thumb, standard operating procedures, practiced behaviors, and selective perception (Kahneman & Tversky, 2000).

What, then, is the utilization heuristic? Basically, the utilization heuristic for managing situational complexity in U-FE is to *stay focused on use*. For every issue that surfaces in evaluation negotiations, for every design decision, for every budget allocation, and for every choice among alternatives, keep asking, “How will this affect use in this situation?” (Patton, 2014).

U-FE FACILITATION PRINCIPLES AND SKILLS

Over the last two decades, the profession of evaluation has been grappling with delineating fundamental and universal evaluation competencies. In 2018, the American Evaluation Association (AEA) adopted a competencies framework with five domains:

professional practice, methodology, context, planning and management, and the interpersonal domain. The interpersonal domain is especially important for U-FE because of the importance of facilitating engagement of primary intended users. This led me to write the book, *Facilitating Evaluation* (Patton, 2020c), centered on the principles and dynamics of U-FE.

The five operating principles for facilitation provide specific guidance for how to practice being a utilization-focused evaluation facilitator.

1. *Be guided by the personal factor.* Customize the evaluation to the participants’ knowledge, interests, motivations, and concerns.
2. *Engage through options.* Deliberating on options and expressing preferences increases participants’ understanding of the implications of making certain choices and increases ownership of the decisions and recommendations that emerge from the process.
3. *Observe, interpret, and adapt.* Monitor group dynamics and progress toward priority purposes to inform pacing, changes in the process, and keeping the group moving toward desired outcomes. Adapt facilitation to what emerges.
4. *Embed evaluative thinking throughout.* This is a defining feature of U-FE facilitation derived from the discovery and conceptualization of process use.
5. *Invigorate with leading-edge inputs.* Evaluation is an evolving field of practice. Part of an evaluation facilitator’s job is to bring before participants new directions, emergent thinking, evolving possibilities, creative opportunities, and inspiring innovations.

U-FE OFFSPRING

Facilitating progress to the leading edge to meet the information needs of primary intended users has led to entirely new evalu-

ation approaches under the umbrella of U-FE.

Developmental Evaluation

In conducting a 5-year U-FE with the Blandin Foundation Community Leadership program, we found that the standard formative–summative evaluation purposes were insufficient to provide ongoing guidance to the program for adaptation and innovation. This program’s intended users (program leadership and staff) needed a more dynamic, real-time feedback approach to support ongoing innovation. This gave rise to a new utilization-focused option: *developmental evaluation* (Patton, 2011; Chapter 21, this volume). How do innovative programs and initiatives adapt within and navigate the turbulence and uncertainties of complex systems change? They adhere to principles.

Principles-Focused Evaluation

Principles-focused evaluation (Patton, 2018) has emerged as a major utilization-focused inquiry framework. Principles constitute a distinct evaluand, a specific and unique focus for evaluation. Principles can guide adaptation. The evaluation question is: What principles work for what situations, with what results? Principles are hypotheses, not truths. They may or may not work. They may or may not be followed. They may or may not lead to desired outcomes. Whether they work, are followed, and yield desired outcomes is subject to evaluation.

Principles-focused evaluation emerged from a utilization-focused meta-evaluation of international development aid principles in 2011/2012 that examined the extent to which aid principles were meaningful, adhered to, and leading to desired development results. The evaluation was awarded the 2012 AEA Outstanding Evaluation Award. Other opportunities to engage in principles-focused evaluation emerged, and it became an identifiable option for

intended use by intended users (see Figure 16.1).

Blue Marble Evaluation

On December 7, 1972, the first photograph of the whole Earth from space was taken by the astronauts in *Apollo 17*. That photo became known as the Blue Marble shot. Blue Marble evaluation takes its name from that whole-Earth image. The Blue Marble perspective means looking beyond nation-state boundaries and across sector and issue silos to connect the global and the local, connect the human and ecological, and connect evaluative thinking and methods with those trying to bring about global systems transformation. Blue Marble evaluation focuses on evaluating the transformations necessary to reverse negative climate change effects and make human life on Earth more sustainable and equitable (Patton, 2020a).

Blue Marble evaluation emerged in 2015, which the United Nations designated at the International Year of Evaluation. It was also the year that the Sustainable Development Goals (SDGs) were adopted by all countries in the world as the pathway for the future of humanity. I was doing extensive training around the world that year on U-FE and developmental evaluation. As I did so, it became increasingly clear that global systems change posed a new set of evaluation challenges, with a great variety of intended users worldwide focused on global systems transformation as the intended use of evaluation. At the same time, I began working with the Global Alliance for the Future of Food (GA), a 30-member collaboration of philanthropic foundations across four continents. What began as U-FE morphed into principles-focused developmental evaluation and ultimately Blue Marble evaluation. The GA created the world’s first Blue Marble evaluation position.

We humans are using our planet’s resources, and polluting and warming it, in ways that are unsustainable. Many people,

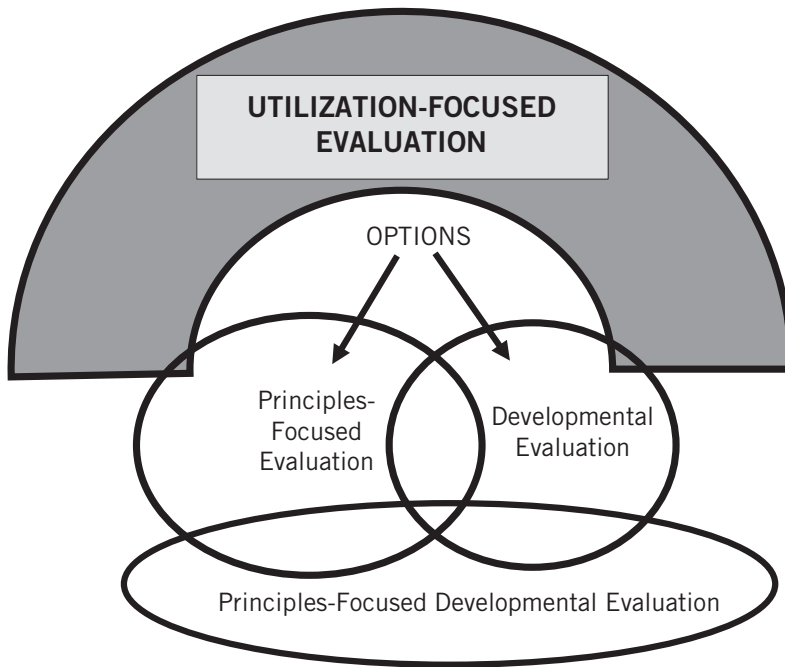


FIGURE 16.1 Relationships among utilization-focused evaluation, developmental evaluation, and principles-focused evaluation. From Patton (2018). Copyright © 2018 The Guilford Press. Reprinted by permission.

organizations, and networks are working to ensure that the future is more sustainable and equitable. Blue Marble evaluators enter the fray by helping design such efforts, providing ongoing feedback for adaptation and enhanced impact, and examining the long-term effectiveness of such interventions and initiatives. Forecasts for the future of humanity run the gamut from doom and gloom to utopia. Evaluation as a transdisciplinary, global profession has much to offer in navigating the risks and opportunities that arise as global change interventions are designed and aimed at a more sustainable and equitable future.

Sustainability and Equity

The theme of the 2019 conference of the International Development Evaluation Association (IDEAS) in Prague was *Evaluation for Transformative Change*, supported by a publication with that title (van den Berg,

Magro, & Mulder, 2019). At the conclusion of the conference, participants from around the world adopted a “Declaration on Evaluation for Transformational Change.” The Declaration included a commitment to focus on sustainability: “In all our evaluations, we commit to evaluating for social, environmental and economic sustainability and transformation, including by assessing contextual factors and systemic changes” (IDEAS, 2019, #6).

Treating environmental sustainability as a universal evaluation criterion in judging the excellence of programs and other interventions is recognition that everyone everywhere is affected ultimately by the global climate emergency.

In 2018, the membership of the AEA approved revised Guiding Principles for Evaluators that includes a principle on Common Good and Equity: Evaluators strive to contribute to the common good and advancement of an equitable and just society.

The Equitable Evaluation Initiative (EEI) and related evaluation initiatives advocate making equity a universal evaluation criterion. This means that all evaluations would include questions about how any particular program, project, policy, intervention, or initiative is addressing equity.

THE U-FE APPROACH TO CRITERIA OF SUSTAINABILITY AND EQUITY

The fifth edition of *Utilization-Focused Evaluation*, published in 2021, makes consideration of the criteria of sustainability and equity a responsibility of utilization-focused evaluators. This constitutes a significant new direction for U-FE that deserves explanation.

Determining what criteria will be used to judge excellence in any evaluation requires the utilization-focused evaluator to facilitate the engagement of primary intended users in determining what criteria are appropriate for a given evaluation. In working with intended users by being active–reactive–interactive–adaptive, the utilization-focused evaluator facilitates users’ selection of criteria for any given evaluation that are situationally and contextually appropriate. That said, here’s an important qualifier to that general approach: In facilitating intended users’ engagement with determining criteria, it is professionally responsible to introduce as items for consideration evaluation criteria that have been widely used and have proven valuable to many, criteria such as the Joint Committee Standards, the AEA Guiding Principles, and the calls for equity and sustainability to be universally incorporated into all evaluations. Intended users need not begin the process of deliberating on principles with a blank slate. Important and thoughtful work has gone into identification of evaluation standards, principles, and criteria. Bringing those forward for consideration is a professional responsibility. Having intended users examine and contemplate sustainability and equity as universal evalu-

ation criteria is how they come to understand the source, importance, relevance, and utility of those criteria. That is how they develop ownership of criteria, including a bit of wordsmithing here and there to make generic and universal criteria situationally applicable and meaningful.

Helping primary intended users understand that evaluators have an obligation to “the common good” means that we do not come to our work with a blank slate where values are concerned. As evaluators, we have a stake in having our evaluations used—and we have a stake in a more equitable and sustainable world. It is appropriate to articulate and own our own stakes and criteria as evaluators. It is just not appropriate to unilaterally impose those on others. Criteria are a core issue for shared negotiation. In making sustainability and equity critical concerns for all evaluations, U-FE joins together the use and values branches of the theory tree (Patton, 2020b).

CONCLUSION

This chapter has traced the evolution of U-FE over nearly five decades. U-FE emerged in 1978 as an outlier perspective arguing the primacy of attention to use when a focus on methodological quality dominated evaluation. Not only has U-FE moved from the margin to the mainstream in evaluation, U-FE has evolved and developed with the evolution and development of the evaluation profession. The discovery of the importance of the personal factor led to defining the focus of U-FE as intended use by intended users. The discovery and conceptualization of process use expanded our understanding of the influence of evaluation beyond just findings to include the learning, capacity development, and engagement in evaluative thinking that occurs through users’ involvement in the evaluation process. More explicit attention to process use led to making more explicit the facilitation skills needed by utilization-

focused evaluators at the same time that the profession adopted competencies that included interpersonal skills. The globalization of evaluation made the profession more sensitive to cultural competences and cross-cultural adaptability, while maintaining a commitment to use as a bedrock of evaluation worldwide.

New evaluation purposes and applications of U-FE have emerged along the way. Working with primary intended users to support their information needs as articulated in their primary intended uses has given rise to entirely new evaluation approaches based on newly appreciated, identified, and conceptualized evaluation uses: developmental evaluation, principles-focused evaluation, and Blue Marble evaluation. While these approaches have led to entire books describing how to conduct evaluations under these innovative frameworks, each one flows from U-FE and each has emerged as an important additional option for consideration by intended users in specifying intended uses.

Now the world is facing a global climate emergency and a crisis of increasing disparities between the wealthy elite and the masses in poverty. The future of humanity is at stake. The profession of evaluation has responded to these deepening crises by adopting sustainability and equity as universal evaluation criteria. U-FE now incorporates those criteria in facilitating the decision making by intended users about what intended uses should be addressed and prioritized.

The next generation of evaluators represented by *EvalYouth*, a global network of young and emerging evaluators, have made a commitment to addressing sustainability and equity as fundamental to evaluation use, thus integrating the use and values branches of the theory tree. My daughter, Charmagne Campbell-Patton, a part of that movement, represents that next-generation perspective as coauthor of the fifth edition of *Utilization-Focused Evaluation* (Patton & Campbell-Patton, 2022). She is among a handful of second-generation evaluators who have taken up the calling of profes-

sional evaluation and, in so doing, are invigorating the profession generally, and U-FE particularly, with new energy, insights, and commitments to making the world a better place through evaluative thinking and engagement, putting theory into practice, and informing further theory developments based on practice. In so doing, they are elevating utilization beyond improving programs and projects to supporting global systems transformation toward a more just and sustainable world.

REFERENCES

- Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568–579.
- Alkin, M. C., & King, J. A. (2017). Definitions and factors associated with evaluation use and misuse. *American Journal of Evaluation*, 38(3), 434–450.
- American Evaluation Association. (2018). *Evaluator competencies*. www.eval.org/p/cm/ld/fid=472
- Becho, L. W. (2019). *An exploratory investigation into the role of evaluation theory in evaluative thinking*. Doctoral dissertation, Western Michigan University. <https://scholarworks.wmich.edu/dissertations/3484>
- Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation*, 40(4), 490–508.
- Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation*, 18(1), 61–77.
- Cousins, J. B. (Ed.). (2020). *Collaborative approaches to evaluation: Principles in use*. SAGE.
- D'Ostie-Racine, L., Dagenais, C., & Ridde, V. (2021). Examining conditions that influence evaluation use within a humanitarian non-governmental organization in Burkina Faso (West Africa). *Systemic Practice Action Research*, 34, 1–35.
- Global Alliance for the Future of Food. (2020). *Home page*. <https://futureoffood.org>
- Haugh, K., & Grodzicki, D. (2016). Bringing the evaluation theory tree to life: #Eval15 Mini-Study [Blog]. <https://aea365.org/blog/katherine-haugh-and-deborah-grodzicki-on->

- bringing-the-evaluation-theory-tree-to-life-eval15-mini-study/?utm_source=newsletter+january+2020&utm_campaign=january+2020+newsletter&utm_medium=email
- International Development Evaluation Association. (2019). *Prague declaration on evaluation for transformational change*. <https://ideas-global.org/wp-content/uploads//2019/10/prague-declaration-4-october-2019.pdf>
- Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge University Press.
- King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, 40(3), 431–458.
- Patton, M. Q. (1978). *Utilization-focused evaluation*. SAGE.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). SAGE.
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). SAGE.
- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4(2), 225–233.
- Patton, M. Q. (2007). Process use as a usefulness. *New Directions for Evaluation*, 116, 99–112.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). SAGE.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.
- Patton, M. Q. (2014). What brain sciences reveal about integrating theory and practice. *American Journal of Evaluation*, 35(2), 237–244.
- Patton, M. Q. (2015). *Essentials of utilization-focused evaluation*. SAGE.
- Patton, M. Q. (2018) *Principles-focused evaluation: The GUIDE*. Guilford Press.
- Patton, M. Q. (2020a). *Blue marble evaluation*. Guilford Press.
- Patton, M. Q. (2020b). Evaluation criteria for evaluating transformation. *American Journal of Evaluation*, 42, 53–89.
- Patton, M. Q. (2020c). *Facilitating evaluation*. SAGE.
- Patton, M. Q., & Campbell-Patton, C. E. (2022). *Utilization-focused evaluation* (5th ed.). SAGE.
- Ramirez, R., & Brodhead, D. (2019). *Comparing the validity of two sets of evaluation principles*. <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/58199/58331.pdf?sequence=1>
- van den Berg, R. D., Magro, C., & Mulder, S. S. (Eds.). (2019). *Evaluation for transformational change: Opportunities and challenges for the sustainable development goals*. International Development Evaluation Association.

CHAPTER 17

Empowerment Evaluation

It Works in Practice, but Does It Work in Theory?

David M. Fetterman

A senior professor and a young graduate student were discussing research at a campus café. The student eagerly shared the results of his work. He could not contain his excitement and described in great detail how well his experiment worked in practice—in a real-world setting. The professor congratulated him on his accomplishment. Then he paused, and with a contemplative and skeptical look on his face, asked the student: But does it work in theory?¹

The jest is designed to portray the out-of-touch, highly abstract thinking associated with academe or the ivory tower. While there is much truth to the sentiment behind this story, it raises a good and valid question. The answer, as outlined in this discussion, is that empowerment evaluation works both in practice and in theory.

AN INTRODUCTION TO EMPOWERMENT EVALUATION

Empowerment evaluation is the use of evaluation concepts, techniques, and findings to

foster improvement and self-determination (Fetterman, 1994). An expanded definition is as follows: “Empowerment evaluation is an evaluation approach that aims to increase the likelihood that programs will achieve results by increasing the capacity of program stakeholders to plan, implement, and evaluate their own programs” (Wandersman et al., 2005, p. 27).

The purpose of empowerment evaluation is empowerment. It is designed to cultivate and enhance self-determination and self-efficacy. Empowerment evaluation helps people help themselves by taking more control over their lives and improving their life trajectories. Empowerment evaluation is typically an ongoing or formative stakeholder involvement approach, in which people take responsibility for monitoring and evaluating their progress, with a focus on results, outcomes, and impact. It can also be summative in nature, helping a group make go/no-go decisions concerning the continuation or elimination of specific programs.

Empowerment evaluation can be used to work with small, medium, or large groups

and communities, ranging from small townships to large-scale Silicon Valley corporations, including Google and Hewlett-Packard. It has been used by tribes on reservations, parents, teachers, and students in school districts, and health care advocates and providers across countries. It is not limited by the type of program, as evidenced by past practice (e.g., elementary education, higher education, hospitals, medical schools, technology-oriented corporations, nonprofits, faith-based organizations, and small businesses; Fetterman, Kaftarian, & Wandersman, 2015).

There are many ways in which to implement an empowerment evaluation. However, one of the most popular approaches is a three-step model.² These steps help to further explicate the nature of empowerment evaluation. These theories, principles, concepts, and steps are interrelated and reinforcing. The theories provide a 30,000-foot view of the approach, while the steps provide a turn-by-turn set of instructions. Together they provide a rich and layered map of the dynamic terrain of empowerment evaluation.

EMPOWERMENT EVALUATION IN PRACTICE

Empowerment evaluation has worked in practice for over two decades. It has been used by the National Aeronautics and Space Administration/Jet Propulsion Laboratory to educate youth about the prototype Mars Rover, in townships in South Africa to create sustainable community health initiatives, by the U.S. Department of Education's Office of Special Education and Rehabilitation Services to foster self-determination, and with Native American tribes to build technological and economic infrastructures on reservations (Fetterman, 2013). Likewise, it has branched out into many additional areas, including schools in academic distress (Fetterman, 2005), accreditation in higher education (Fetterman, 2011), minority tobacco prevention (Fetterman, Delaney,

Triana-Tremain, & Evans-Lee, 2015), and medical education (Fetterman, 2009; Fetterman, Deitz, & Gesundheit, 2010). It has been used by Peruvian women artisans to build sustainable businesses (Sastre-Merino, Vidueira, Diaz-Puente, & Fernande-Moral, 2015), teachers to evaluate their own performance (Clinton & Hattie, 2015), and fourth and fifth graders to enhance diversity and inclusion (Langhout & Fernandez, 2015). Recently, it has been used to facilitate evaluations remotely, including Addiction Studies Programs and tuberculosis elimination initiatives in India (sponsored by the U.S. Agency for International Development).

"Nothing Is as Practical as a Good Theory" (Empowerment Evaluation in Theory)

Kurt Lewin is credited with coining the phrase: "Nothing is as practical as a good theory" (see Lewin, 1951; Greenwood & Levin, 1998, p. 19). Good theory guides effective action. It provides the evaluator with a road map of the journey with useful, accurate, consistent, and logical steps. My father used to say that "if you don't know where you are going, any road will get you there." A good theory, like a good map, can minimize missteps and increase the likelihood of reaching your destination.

Theories guiding empowerment evaluation, as noted earlier, include empowerment and self-determination theories. It is also informed by specific evaluation theories, including process use and theories of use and action. In turn, these theories help inform 10 overarching principles that provide empowerment evaluation with an explicit direction and purpose, beginning with improvement and continuing to accountability. Key concepts that help define empowerment evaluation include critical friends, cultures of evidence, cycles of reflection and action, communities of learners, and reflective practitioners.

These are sequentially ordered conceptual building blocks. Once defined, they begin with the macrolevel or highest level of

abstraction, theories, and end at the micro-level, the specific steps of empowerment evaluation. This serves to place the steps in a theoretical and conceptual context, much the same as other theories, models, and concepts are used to place the steps of alternative approaches in context. The sequence is designed to help practitioners understand and implement empowerment evaluation practice. In addition, it is necessary to describe the approach in order to test whether it is being implemented with fidelity.

Theories

The theories behind this simple process are surprisingly complex. These theories are the underpinnings of this approach. An exploration into the theories guiding empowerment practice helps to illuminate the integral relationship between method and use in empowerment evaluation. The most pertinent theories guiding empowerment evaluation are empowerment theory, self-determination theory, process use, and theories of use and action. *Empowerment theory* is divided into processes and outcomes. This theory has implications for the role of the empowerment evaluator or facilitator, which differs from a traditional evaluator. *Self-determination* is one of the foundational concepts underlying empowerment theory and it helps to detail specific mechanisms or behaviors that enable the actualization of empowerment. *Process use* represents much of the rationale or logic underlying empowerment evaluation in practice because it cultivates ownership by placing the approach in community and staff members' hands. Finally, the alignment of *theories of use and action* explains how empowerment evaluation helps people produce desired results.

Empowerment Theory

Writings in the community psychology have developed an area called empowerment theory, which is about gaining control, obtain-

ing resources, and understanding one's social environment. It is also about problem solving, leadership, and decision making. It operates on many levels, and distinguishing between empowering processes and outcomes is critical. According to Zimmerman (2000, pp. 44–45):

Empowerment processes are ones in which attempts to gain control, obtain needed resources, and critically understand one's social environment are fundamental. The process is empowering if it helps people develop skills so they can become independent problem solvers and decision makers. Empowering processes will vary across levels of analysis. For example, empowering processes for individuals might include organizational or community involvement, empowering processes at the organizational level might include shared leadership and decision making, and empowering processes at the community level might include accessible government, media, and other community resources.

Similar to the distinctions of process and outcome in evaluation, empowerment theory processes contribute to specific outcomes. Linking the processes to outcomes helps draw metalevel causal relationships or at least a chain of reasoning. When specified outcomes are achieved, it is possible to retrace steps to determine which processes were most effective. Similarly, when specific processes are implemented poorly, it is easy to see the causal relationship or at least the contributing factors associated with a failure to achieve specified outcomes. Zimmerman (2000, pp. 44–45) provides additional insight into the outcome level of analysis to further explicate empowerment theory:

Empowerment outcomes refer to the operationalization of empowerment so we can study the consequences of citizen attempts to gain greater control in their community or the effects of interventions designed to empower participants. Empowered outcomes also differ across levels of analysis. When we are concerned with individuals, outcomes might include situation-specific perceived control,

skills, and proactive behaviors. When we are studying organizations, outcomes might include organizational networks, effective resource acquisition, and policy leverage. When we are concerned with community-level empowerment, outcomes might include evidence of pluralism, the existence of organizational coalitions, and accessible community resources.

Role

Zimmerman's (2000) characterization of the community psychologist's *role* in empowerment activities is easily adapted to the empowerment evaluator. It also demonstrates the impact of theory on practice, shaping every dimension of the approach, including the evaluator's role.

An empowerment approach to intervention design, implementation, and evaluation redefines the professional's role relationship with the target population. The professional's role becomes one of collaborator and facilitator rather than expert and counselor. As collaborators, professionals learn about the participants through their cultures, their worldviews, and their life struggles. The professional works *with* participants instead of advocating *for* them. The professional's skills, interests, or plans are not imposed on the community; rather, professionals become a resource for a community. This role relationship suggests that what professionals do will depend on the particular place and people with whom they are working, rather than on the technologies that are predetermined to be applied in all situations. (pp. 44–45, emphasis in original)

An eloquent literature on empowerment theory by Zimmerman (2000), Zimmerman, Israel, Schulz, and Checkoway (1992), Zimmerman and Rappaport (1988), and Dunst, Trivette, and LaPointe (1992) also informs empowerment evaluation.

*Self-Determination*³

The theoretical level of empowerment processes and outcomes requires mechanisms

to link it to action. Dennis Mithaug's (1991, 1993) extensive work with individuals with disabilities to explore the concept of self-determination provided additional theoretical inspiration and guidance. The concept of self-determination details specific mechanisms that help program staff members and participants implement an empowerment evaluation. Self-determination was one of the foundational concepts in the study that informs empowerment evaluation today. *Self-determination* is defined as the ability to chart one's own course in life. It consists of numerous interconnected capabilities, such as the ability to identify and express needs; establish goals or expectations and a plan of action to achieve them; identify resources; make rational choices from various alternative courses of action; take appropriate steps to pursue objectives; evaluate short- and long-term results, including reassessing plans and expectations and taking necessary detours; and persist in the pursuit of those goals. A breakdown at any juncture of this network of capabilities—as well as various environmental factors—can reduce a person's likelihood of being self-determined.⁴

Dennis and I, as part of an American Institutes for Research team, completed a 2-year grant funded by the Department of Education on self-determination and individuals with disabilities. We conducted research designed to help both providers for students with disabilities and the students themselves become more empowered. We learned about self-determined behavior and attitudes and environmentally related features of self-determination by listening to self-determined children with disabilities and their providers.

Process Use

One of the most significant problems facing the field of evaluation is inadequate knowledge utilization. Evaluation reports often sit and gather dust. Decisions are made without the benefit of being informed by evaluation findings. Empowerment evaluation

is designed to be used by people. It places evaluation in the hands of community and staff members. The more that people are engaged in conducting their own evaluations, the more likely they are to believe in them because the evaluation findings are theirs. In addition, a by-product of this experience is that they learn to think evaluatively. This makes them more likely to make decisions and take actions based on their evaluation data. This way of thinking is at the heart of process use.⁵

According to Patton (2013, p. 301, original emphasis), “Helping people learn to think evaluatively by participating in real evaluation exercises is what I’ve come to call ‘process use’ (Patton 1997a, 1997b). I have defined process use as relating to and being indicated by individual changes in thinking and behaving that occur among those involved in evaluation as a result of the learning that occurs *during the evaluation process*.” In an empowerment evaluation, thinking evaluatively is a product of guided immersion. This occurs when people conduct their own evaluation as guided by an empowerment evaluator. Teaching people to think evaluatively is like teaching them to fish. It can last a lifetime and is what evaluative sustainability is all about — internalizing evaluation (individually and institutionally).

Empowerment evaluation invites, if not demands, participation. Participation or immersion is a form of experiential education. Guided, this immersion helps people to see the world through an evaluative lens. Participation also creates an authentic, credible, and almost palpable sense of ownership. It is this combination of evaluative thought and ownership, through immersion, that makes empowerment evaluation work, improving knowledge utilization in the process.

Theories of Use and Action

Once the groundwork is laid with empowerment, self-determination theory, and pro-

cess use theories, conceptual mechanisms become more meaningful. Theories that enable comparisons between use and action are essential. The approach works when the pieces are in place. When things go wrong, which is normal in life, it is possible to compare and identify the areas needing attention.

Empowerment evaluation relies on the reciprocal relationship between theories of action and use at every step in the process. A *theory of action* is usually the espoused operating theory about how a program or organization works. It is a useful tool, generally based on program personnel views. This theory of action is often compared with a theory of use. The theory of use is the actual program reality, the observable behavior of stakeholders (see Argyris & Schon, 1978; Patton, 1997b). People engaged in empowerment evaluations create a theory of action at one stage and test it against the existing theory of use during a later stage. Similarly, they create a new theory of action as they plan for the future. Because empowerment evaluation is an ongoing and iterative process, stakeholders test their theories of action against theories of use during various microcycles to determine whether their strategies are being implemented as recommended or designed. The theories go hand in hand in empowerment evaluation.

These theories are used to identify gross differences between the ideal and the real. For example, communities of empowerment evaluation practice compare their theory of action with their theory of use to determine whether they are even pointing in the same direction. Three common patterns that emerge from this comparison include *in alignment*, *out of alignment*, and *alignment in conflict* (see Figure 17.1). In alignment is when the two theories are parallel or pointed in the same direction. They may be distant or close levels of alignment, but they are on the same general track. Out of alignment occurs when actual practice is divergent from the espoused theory of how things are supposed to work. The theory of

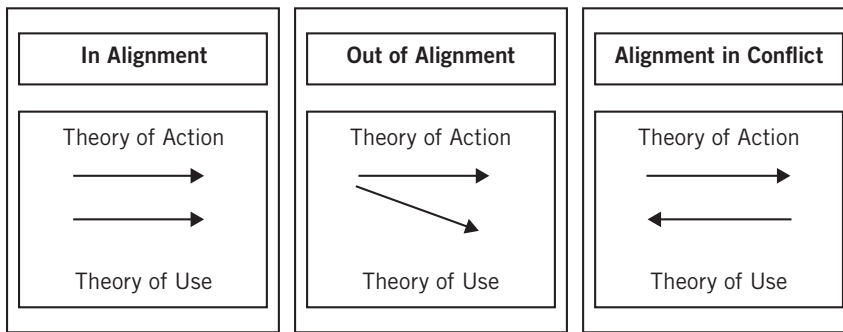


FIGURE 17.1 Contrasting patterns of alignment. From Fetterman and Eiler (2001). Copyright © 2001 by the authors. Reprinted by permission.

use is not simply distant or closely aligned; it is actually off-target or at least pointed in another direction. Alignment in conflict occurs when the theory of action and use are pointed in diametrically opposite directions. This signals a group or organization in serious trouble or self-denial.

After making the first-level comparison, a gross indicator, to determine whether the theories of action and use are even remotely related to each other, communities of empowerment evaluation practice compare their theory of action with their theory of use in an effort to reduce the gap between them. This assumes that they are at least pointed in the same direction. The ideal progression is from distant alignment to close alignment between the two theories. This is the conceptual space in which most communities of

empowerment evaluation practice strive to accomplish their goals as they close the gap between the theories (see Figure 17.2).

The process of empowerment embraces the tension between the two types of theories and offers a means for reconciling incongruities. This dialectic in which theories of action and use are routinely juxtaposed in daily practice creates a culture of learning and evaluation.

Principles

The foundations of empowerment evaluation, theoretically, lead to specific principles to instruct practice. Empowerment evaluation principles provide a sense of direction and purposefulness throughout an evaluation. Empowerment evaluation is guided

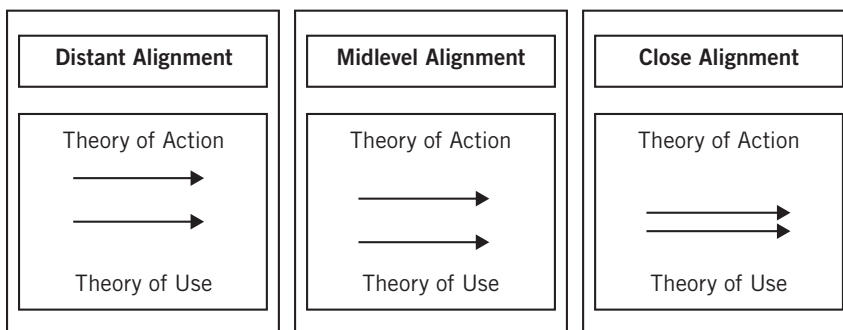


FIGURE 17.2 Aligning theories of action and use to reduce the gap. From Fetterman and Eiler (2001). Copyright © 2001 by the authors. Reprinted by permission.

by 10 specific principles (Fetterman, 2015, pp. 27–28; Fetterman & Wandersman, 2005, pp. 1–2, 27–41, 42–72). They include the following:

1. *Improvement.* Empowerment evaluation is designed to help people improve program performance; it is designed to help people build on their successes and reevaluate areas meriting attention.
2. *Community ownership.* Empowerment evaluation values and facilitates community control; use and sustainability are dependent on a sense of ownership.
3. *Inclusion.* Empowerment evaluation invites involvement, participation, and diversity; contributions come from all levels and walks of life.
4. *Democratic participation.* Participation and decision making should be open and fair.
5. *Social justice.* Evaluation can and should be used to address social inequities in society.
6. *Community knowledge.* Empowerment evaluation respects and values community knowledge.
7. *Evidence-based strategies.* Empowerment evaluation respects and uses the knowledge base of scholars (in conjunction with community knowledge).
8. *Capacity building.* Empowerment evaluation is designed to enhance stakeholders' ability to conduct evaluation and to improve program planning and implementation.
9. *Organizational learning.* Data should be used to evaluate new practices, inform decision making, and implement program practices; empowerment evaluation is used to help organizations learn from their experience (building on successes, learning from mistakes, and making midcourse corrections).
10. *Accountability.* Empowerment evaluation is focused on outcomes and accountability; empowerment evaluations function within the context of existing policies, standards, and measures of account-

ability; did the program accomplish its objectives?

Empowerment evaluation principles help evaluators and community members make decisions that are in alignment with the larger purpose or goals associated with capacity building and self-determination. The principles of inclusion, for example, remind evaluators and community members to include, rather than exclude, members of the community, even though fiscal, logistic, and personality factors might suggest otherwise. The capacity-building principle reminds the evaluator to provide community members with the opportunity to collect their own data, even though it might initially be faster and easier for the evaluator to collect the same information. The accountability principle guides community members to hold one another accountable. It also situates the evaluation within the context of external requirements and credible results or outcomes (see Fetterman, 2005, p. 2).

Concepts

Empowerment evaluation concepts provide a more instrumental view of how to implement the approach. Key concepts include critical friends, cultures of evidence, cycles of reflection and action, communities of learners, and reflective practitioners.⁶ A critical friend is an evaluator who facilitates the process and steps of empowerment evaluation. They believe in the purpose of the program but provide constructive feedback. They help to ensure the evaluation remains organized, rigorous, and honest.

Empowerment evaluators help cultivate a culture of evidence by asking people why they believe what they believe. They are asked for evidence or documentation at every stage so that it becomes normal and expected to have data to support one's opinions and views. Cycles of reflection and action involve ongoing phases of analysis, decision making, and implemen-

tation (based on evaluation findings). It is a cyclical process. Programs are dynamic, not static, and require continual feedback as they change and evolve. Empowerment evaluation is successful when it is institutionalized and becomes a normal part of the planning and management of the program.

Empowerment evaluation is driven by a group process. It is a community of learners. The group members learn from each other, serving as their own peer review group, critical friend, resource, and norming mechanism. Individual members of the group hold each other accountable concerning progress toward stated goals. Finally, empowerment evaluations help create reflective practitioners. Reflective practitioners use data to inform their decisions and actions concerning their own daily activities. This produces self-aware and self-actualized individuals who have the capacity to apply this worldview to all aspects of their lives. As individuals develop and enhance their own capacity, they improve the quality of the group's exchange, deliberation, and action plans.

Steps

There are many ways in which to implement an empowerment evaluation. In fact, empowerment evaluation has accumulated a warehouse of useful tools. The three-step approach to empowerment evaluation is one of the most popular tools in the collection (Fetterman, 2001). It includes helping group members (1) establish their mission, (2) take stock of their current status, and (3) plan for the future. The popularity of this particular approach is in part a result of its simplicity, effectiveness, and transparency.

Mission

The group members come to a consensus concerning their mission or values (see Figure 17.3). This gives them a shared vision of what's important to them and where they want to go. The empowerment evaluator facilitates this process by asking participants



FIGURE 17.3 Mission.

to generate statements that reflect their mission. These phrases are recorded on a poster sheet of paper (and may be projected on a liquid crystal display [LCD] projector depending on the technology available). Alternatively, participants can collaborate remotely with the use of various online programs. These phrases are used to draft a mission statement (crafted by a member of the group and the empowerment evaluator). The draft is circulated among the group members. They are asked to “approve” it and/or suggest specific changes in wording as needed. A consensus about the mission statement helps the group members think clearly about their self-assessment and plans for the future. It anchors the group in common values.

Taking Stock

After coming to a consensus about the mission, the group members evaluate their efforts (within the context of a set of shared

values). First, the empowerment evaluator helps members of the group generate a list of the most important activities required to accomplish organizational or programmatic goals. The empowerment evaluator gives each participant five dot stickers and asks the participants to place them by the activities they think are the most important to accomplish programmatic and organizational goals (and thus the most important to evaluate as a group from that point on). They can put one sticker on five different activities or all five on one activity if they are concerned that activity will not get enough votes (see Figure 17.4). Once again, this is also being conducted online as needed. The top 10 items with the most dots represent the results of the prioritization part of taking stock. The 10 activities represent the heart of the second part of taking stock: rating.

The empowerment evaluator asks participants in the group to rate how well they are doing concerning each of the activities



FIGURE 17.4 Taking Stock: Part I.



FIGURE 17.5 Taking Stock: Part II.

selected, using a 1 (*low*) to 10 (*high*) scale (see Figure 17.5). The columns are averaged horizontally and vertically. Vertically, the group can see who is typically optimistic and/or pessimistic. This helps the group calibrate or evaluate the ratings and opinions of each individual member. It helps the group establish norms. Horizontally, the averages provide the group with a consolidated view of how well (or poorly) things are going. The empowerment evaluator facilitates a discussion and dialogue about the ratings, asking participants why they gave a certain activity a 3 or a 7.

The dialogue about the ratings is one of the most important parts of the process. In addition to clarifying issues, evidence is used to support viewpoints, and “sacred cows” are surfaced and examined during dialogue. Moreover, the process of specifying the reason or evidence for a rating provides the group with a more efficient and focused manner of identifying what needs

to be done next, during the planning for the future step of the process. Instead of generating an unwieldy list of strategies and solutions that may or may not be relevant to the issues at hand, the group can focus its energies on the specific concerns and reasons for a low rating that were raised in the dialogue or exchange.

Planning for the Future

Many evaluations conclude at the taking stock phase. However, taking stock is a baseline and a launching off point for the rest of the evaluation. After rating and discussing programmatic activities, it is important to do something about the findings. It is time to plan for the future (see Figure 17.6). This step involves generating goals, strategies, and credible evidence (to determine if the strategies are being implemented and if they are effective). The goals are directly

related to the activities selected in the taking stock step. For example, if communication was selected, rated, and discussed, then communication (or improving communication) should be one of the goals. The strategies emerge from the taking stock discussion as well, as noted earlier. For example, if communication received a low rating and one of the reasons was because the group members never had agendas for their meetings, then preparing agendas might become a recommended strategy in the planning for the future exercise.

Monitoring the Strategies

Many programs, projects, and evaluations fail at this stage for lack of individual and group accountability. Individuals, who spoke eloquently and/or emotionally about a certain topic should be asked to volunteer to lead specific task forces to respond to identified problems or concerns (see Figure 17.7). They do not have to complete the task. However, they are responsible for taking the lead in a circumscribed area (a specific goal) and reporting the status of the effort periodically at ongoing management meetings. Similarly, the members should make a commitment to reviewing the status of these new strategies as a group (and be willing to make midcourse corrections if they are not working). Conventional and innovative evaluation tools are used to monitor the strategies, including online surveys, focus groups, interviews, as well as the use of a quasi-experimental design (if appropriate). In addition, program-specific metrics are developed using baselines, milestones, and goals (as deemed useful and appropriate). For example, a minority tobacco prevention program empowerment evaluation in Arkansas has established the following:

1. Baselines (the number of people using tobacco in their community)
2. Goals (the number of people they plan to help stop using tobacco by the end of the year)



FIGURE 17.6 Planning for the Future.



FIGURE 17.7 Comparing baselines with benchmarks and goals.

3. Milestones (the number of people they expect to help stop using tobacco each month)
4. Actual performance (the number of people they help stop smoking)

The actual performance is compared with the milestones and goals in order to determine if the group is making progress. These metrics are used to help a community monitor program implementation efforts

and enable program staff and community members to make midcourse corrections and substitute ineffective strategies for potentially more effective ones as needed. The metrics are not used to chastise, stigmatize, or ostracize individuals. They simply serve to signal when the group needs assistance. These data are also invaluable when the group conducts a second taking stock exercise (3 to 6 months later) to determine if they are making progress toward their desired goals and objectives.

THEORY IN PRACTICE: A USEFUL GUIDE

The essence of empowerment evaluation is a systematic way of thinking. Empowerment evaluation is a constellation of these theories, concepts, principles, and steps. They are used to guide practice. Empowerment evaluation is more than the sum of its parts (including “essential” parts, discussed later in this text). It is the gestalt or whole package that makes it work. Together, empowerment evaluation ideas, values, and practices help people learn how to think like an evaluator. They build their evaluation capacity in the process of evaluating the impact of their own work. This approach to evaluation fosters improvement and self-determination.

Process Use

The theory of process use guides daily practice. The evaluation is purposely placed in the hands of the program staff and community members. They pose the evaluation questions (within the context of their preestablished commitments to funders and community members). They collect the data, analyze, make meaning, and report learnings (with the assistance of a critical friend or empowerment evaluator). This level of stakeholder involvement creates a sense of ownership, which in turn enhances credibility, follow-through, and sustainability. They find it credible because they create, conduct, and own the evaluation.

Theories of Use and Action

The continual juxtaposition of theories of use and action helps community and staff members test the degree of alignment between what their organization or community espouses and does in practice. It helps people think like evaluators.

Principles

The 10 principles of empowerment evaluation work together synergistically. For example, the first principle, improvement, reflects the pragmatic and utilitarian nature of empowerment evaluation. The aim is to help people improve their programs and practice and succeed in accomplishing their objectives. Community ownership is required to make this happen in a meaningful and sustained manner. Authentic community ownership requires inclusion. It cannot be a single elite group making all the decisions. People from all parts of the organization and/or community should be to be included. Participation from many stakeholders, including those typically marginalized or excluded, is critical if the effort is to be credible and taken seriously. It is also more efficient to include major stakeholders at the beginning rather than having to revisit each of the issues every time a new group is invited to participate in the group.

It is that interconnected, interrelated, and reinforcing nature of empowerment evaluation that gives it strength and sustainability. The same type of synergy and interconnectivity applies to the remaining combination of principles (see Fetterman & Wandersman, 2005, pp. 210–212 for details). The interaction among the participants and the principles results in a rising level of empowerment and self-determination (see Fetterman & Wandersman, 2005, p. 213).

However, empowerment evaluation is not a one-size-fits-all approach. People adapt it to local circumstances, conditions, and needs. It is designed to help people conscientiously assess the impact of their work,

which is shaped, not driven, by local circumstances, conditions and needs.

Three Examples

Three empowerment evaluation principles have been selected to help highlight the holistic nature of empowerment evaluation. Ownership, capacity building, and accountability principles are important threads of empowerment evaluation. A brief description of these principles demonstrates how, together, they weave a way of thinking and acting into the tapestry of empowerment evaluation. (These three have been selected for illustration purposes. This example is not designed to minimize the need to apply as many principles as possible in a given effort.)

Community Ownership

People have the right to make decisions about actions that affect their lives. Putting evaluation in the hands of program staff and participants is thought to foster self-determination and responsibility instead of dependency. In addition, as stated earlier, empowerment evaluation is guided by the theory (process use) that people are more likely to believe and use findings and follow recommendations if they were responsible for creating them.

Capacity Building

Empowerment evaluation's emphasis on capacity building and process use helps people think like evaluators. Evaluation capacity building is related to producing outcomes (see Labin, Duffy, Meyers, Wandersman, & Lesesne, 2013). If there is no capacity building, then there is no empowerment evaluation.

Accountability

Empowerment evaluation has always been committed to producing outcomes. It is not

simply an engaging process. Empowerment evaluators, funders, staff members, and community members share a common commitment to producing results. The list of real-world outcomes associated with empowerment evaluations is both long and significant (see Fetterman et al., 2015).

“Nothing as Theoretical as Good Practice”

A solid argument can be made for the role of theory guiding practice. However, consideration should also be given the complete opposite proposition. There is also “nothing as theoretical as good practice.”

According to Ployhart and Bartunek (2019, p. 493), editors of the *Academy of Management Review* (AMR), who recognized that theoretical contributions are often not actionable and may, in fact, do more harm than good, “the vast majority of theoretical developments in AMR . . . offer implications for practice that may or may not be actionable (Bartunek & Rynes, 2010). At worst, they may destroy good organizational practice (Ghoshal, 2005).”

Glaser and Strauss (in Bryant & Charmaz, 2010) have proposed a “grounded theory” approach that derives theory from day-to-day, empirical practice. Empowerment evaluation is largely informed from the insider’s or emic perspective and grounded theory. In addition to helping construct theory, empowerment evaluation theory is continually tested by actual behavior. In empowerment evaluation, if the theory does not match the behavior, then the theory is adjusted, not the behavior.

Ployhart and Bartunek (2019) conclude their observations: “Perhaps, rather than assuming that we as academics have a few precious bits to bestow on practitioners, we should consider that we would not be able to do our work without theirs. So, with thanks to Lewin, we suggest that it’s time to turn his famous maxim on its head. There is nothing so theoretical as good practice” (p. 493). The take-home from this exercise in linguistic gymnastics is that theory and prac-

tice in empowerment evaluation are equally instructive to learning.

TRANSLATING THEORY INTO PRACTICE AND PRACTICE INTO THEORY (FOCUSING ON ESSENTIAL FEATURES)

Empowerment evaluation’s essential features include the theory of process use, theory of action, and theory of use. It also includes key concepts, guiding principles, and specific steps. Highlighting a sample of these theories, principles, and steps provides evaluators, sponsors, and community members with guidance as they verify fidelity to a model or high, medium, and low levels of empowerment evaluation implementation (see Fetterman & Wandersman [2005] for levels of empowerment evaluation).

Theory

Process use is an essential feature of empowerment evaluation. It represents much of the rationale or logic underlying empowerment evaluation in practice because it cultivates ownership by placing the approach in community and staff members’ hands. The more that people are engaged in conducting their own evaluations, the more likely they are to believe in them because the evaluation findings are their own. This makes them more likely to make decisions and take actions based on their evaluation data.

Once the groundwork is laid with empowerment and process use theories, conceptual mechanisms become more meaningful. Theories that enable comparisons between use and action are also essential. People engaged in empowerment evaluation create a theory of action at one stage and test it against the existing theory of use during a later stage. It helps people determine consistencies and inconsistencies in organizational and community behavior. A group can identify where and when it is not “walking its talk.” This dialectic in which theories of action and use are routinely jux-

taposed in daily practice creates a culture of learning and evaluation that is essential to empowerment evaluation.

Principles

The 10 empowerment evaluation principles provide a sense of direction and purposefulness throughout an evaluation. Core principles include improvement, community ownership, inclusion, capacity building, organizational learning, and accountability. Empowerment evaluation principles help evaluators and community members make decisions in alignment with empowerment evaluation goals associated with capacity building and self-determination. Although each principle is important, the community members are in the best position to highlight the principles that best match their needs. In other words, in the ideal world, all of the principles are selected, but in the real world, a sample of the most appropriate principles from the list are used to guide a specific empowerment evaluation. They are dependent on community context and capacity.

A Critical Friend

A critical friend is one of the most important roles played in an empowerment evaluation. It is difficult to imagine how an empowerment evaluation could function without one. The role of the critical friend merits attention because it is like a fulcrum in terms of fundamental relationships. Applied improperly, it can be like a wedge inhibiting movement and change; applied correctly, this role can be used to leverage and maximize the potential of a group. Some people ask, How can an empowerment evaluator be objective and critical if they are friends and in favor of a type of program? The answer is simple: An empowerment evaluator is critical and objective because he or she wants the program to work (or work better). He or she may be in favor of a general type of program but do not assume a position about a specific program without data. Important attri-

butes of a critical friend include (1) creating an environment conducive to dialogue and discussion, (2) providing or requesting data to inform decision making, (3) facilitating rather than leading, (4) being open to ideas and inclusive, and (5) being willing to learn (see Fetterman, 2009; Fetterman, Deitz, & Gesundheit, 2010).

Empowerment evaluators help cultivate a culture of evidence by asking people why they believe what they believe. They are asked for evidence or documentation at every stage so that it becomes normal and expected to have data to support one's opinions and views. Empowerment evaluators are essential because they help ground the group in data and data-driven decision making.

Steps

There are many ways to conduct an empowerment evaluation. The three-step approach, as discussed earlier, includes helping group members (1) establish their mission, (2) take stock of their current status, and (3) plan for the future. The popularity of this particular approach is in part a result of its simplicity, effectiveness, and transparency. The steps provide people with a logic model that, like a highway, they build while driving on it. Their "Plans for the Future" or intervention is rooted in a self-assessment (or Taking Stock exercise), and the self-assessment is rooted in their mission. This creates a common thread or chain of reasoning. In addition, focusing on the mission or "big picture," refining the group's efforts by prioritizing the most important activities to monitor and evaluate as a group, and then implementing "Plans for the Future" (based on Taking Stock assessments and exchange) produces a greater "dose effect"—focusing the group's efforts from a large-scale, "big picture" mission to a small list of critical activities.

Adopting and applying a three-step or a 10-step approach makes it easy to certify that an empowerment evaluation approach is being adopted (within the context of

the theories, concepts, and principles presented). However, they are not the only steps used to facilitate an empowerment evaluation. Stories, maps of the “journey,” and other graphics have been used effectively. The specific steps selected to facilitate an empowerment evaluation are only limited by our creativity and imagination. However, the steps selected should model the logic and values associated with these two popular and effective approaches if they are to conform with some meaningful fidelity to the model.

Tools

The evaluation dashboard helps build capacity by providing people with a tool they create to help them monitor their own performance over time. The dashboard is shared with the evaluation team, administrators, and sponsors. This eliminates the need to “game the system,” engage in subterfuge, or make excuses. Instead of being shamed, blamed, or slapped on the wrist for not reaching milestones, the dashboard is used to signal the need for help, support, and guidance. It also helps community members learn how to determine when it is necessary to make midcourse corrections. Ultimately, this tool is designed to help people stay on track concerning the desired goals and bring support when needed (before it is too late). This positive, nurturing, and constructive approach is essential to a healthy and productive form of empowerment evaluation.

These essential features provide an insight into elements required to determine if there is fidelity to the theoretical model. This list of essential features is not meant to be exhaustive, but without adherence to the essential features described, it is unlikely that an empowerment evaluation approach has been adopted.

CONCLUSION

Many of the theories guiding empowerment evaluation have emerged from practice. In

the true spirit of reflexivity, many empowerment evaluation theories are now used to guide practice. I began this discussion by asking if empowerment evaluation works in theory. Ironically, it does, but the test of any good empowerment evaluation theory is also to ask whether it works in practice.

Stanford University School of Medicine’s empowerment evaluation is a useful example. They used an empowerment evaluation approach to inform curricular decision making (Fetterman et al., 2010). The process was placed in the hands of faculty, students, staff members, and administrators in order to help them participate in system changes. Patients were also involved in the courses to enhance the clinical dimension.

Empowerment evaluation provided them with a method for gathering, analyzing, and sharing data about their academic program and its outcomes. Process use, as applied to this example, assumed that the more closely stakeholders are involved in conducting their own evaluations and reflecting on their data, the more likely they are to take ownership of the results and to guide curricular decision making and reform. Students, faculty, and administrators participated in large numbers, in a substantive manner, and over a prolonged period. They embraced the process, data, and decision making.

Another test of the approach focuses on measuring outcomes. What were the results? Did it have any impact? Applying process use to curriculum evaluation at Stanford contributed to improvements in course and clerkship ratings. In comparing evaluation results before and after stakeholders began using this approach (and in essence took control of their own evaluation), we found that the average student ratings for required courses improved significantly ($P = .04$; student’s one-sample *t*-test).

The use of empowerment evaluation, and the totality of theories, concepts, principles, and steps, fostered greater institutional self-reflection, led to an evidence-based model of decision making, and expanded opportunities for students, faculty, and support staff to work collaboratively to improve and

refine the medical school curriculum. It was also instrumental in helping the medical school remain accredited.

Similar examples are drawn from Hewlett-Packard's empowerment evaluation of their Digital Villages (designed to bridge the digital divide in communities of color). The Tribal Digital Village used empowerment evaluation to guide their efforts in building one of the largest unlicensed wireless systems in the country. The list of accomplishments and outcomes is extensive and global in scope.

The journey from theory to practice and back again in empowerment evaluation has been long and often difficult but productive. We appreciate how far we have come and Patton's placement of empowerment evaluation in the "pantheon of major approaches" (2015, p. x). We are also sincerely humbled by Donaldson's (2015, p. ix) assessment of our contributions: "One of the greatest evaluation innovations of the past two decades has been the development of professional and systematic approach to self-evaluation called empowerment evaluation."

Nevertheless, we recognize how much more work lies ahead as we continue to enhance conceptual clarity and methodological specificity. We live in a world where, without warning, the entire planet and our continued existence as a species is called into question by global warming, social inequities, nuclear threats, and even a microscopic virus. These challenges only strengthen our resolve to continue to learn how to adapt this global phenomenon to a world that is, in theory, in a constant state of change.

NOTES

1. The earliest recording of this joke was in 1911, in an issue of a Boston-based magazine titled *The Youth's Companion: For All the Family*.
2. There is also an equally popular 10-step approach referred to as Getting to Outcomes (2004), developed by Chinman, Imm, and Wandersman.
3. Mithaug's work in this area is typically

referred to as self-regulation theory, with self-determination as a guiding concept in his work. For the purposes of clarity and as it relates to instructing empowerment evaluation, self-determination is being used as the umbrella term in this discussion.

4. See also Bandura (1982) for more detail on issues related to self-efficacy and self-determination.
5. There is a substantial literature concerning the use of evaluation. However, most of it is devoted to lessons learned after the evaluation. The discussion of process use in this context focuses on use during an evaluation.
6. These concepts are influenced by traditional organizational development and transformation theorists including Argyris and Schon (1978) and Senge (1994), as well as evaluators associated with organizational learning (Preskill & Torres, 1999).

REFERENCES

- Argyris, C., & Schon, D. A. (1978). *Organizational learning: A theory of action perspective*. Addison-Wesley.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122–147.
- Bartunek, J. M., & Rynes, S. L. 2010. The construction and contributions of "implications for practice": What's in them and what might they offer? *Academy of Management Learning & Education*, 9, 100–117.
- Bryant, A., & Charmaz, K. (2010). *The SAGE handbook of grounded theory*. SAGE.
- Chinman, M. Imm, P., & Wandersman, A. (2004). *Getting to Outcomes: Promoting accountability through methods and tools for planning, implementation, and evaluation*. RAND Corporation.
- Clinton, J., & Hattie, J. (2015). Teachers as evaluators: An empowerment evaluation approach. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Donaldson, S. (2015). Foreword. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Dunst, C. J., Trivette, C. M., & LaPointe, N.

- (1992). Toward clarification of the meaning and key elements of empowerment. *Family Science Review*, 5, 111–130.
- Fetterman, D. M. (1994). Empowerment evaluation. *Evaluation Practice*, 15(1), 1–15.
- Fetterman, D. M. (2001). *Foundations of empowerment evaluation*. SAGE.
- Fetterman, D. M. (2005). Empowerment evaluation: From the digital divide to academic distress. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice*. Guilford Press.
- Fetterman, D. M. (2009). Empowerment evaluation at the Stanford University School of Medicine: Using a critical friend to improve the clerkship experience. *Ensaio: Avaliação e Políticas Públicas em Educação, Rio de Janeiro*, 17(63), 197–204.
- Fetterman, D. M. (2011). Empowerment evaluation and accreditation case examples: California Institute of Integral Studies and Stanford University. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education*. Routledge.
- Fetterman, D. M. (2013). *Empowerment evaluation in the digital villages: Hewlett-Packard's \$15 million race toward social justice*. Stanford University Press.
- Fetterman, D. M. (2015). Empowerment evaluation: Theories, principles, concepts, and steps. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Fetterman, D. M., Deitz, J., & Gesundheit, N. (2010). Empowerment evaluation: A collaborative approach to evaluating and transforming a medical school curriculum. *Academic Medicine*, 85(5), 813–820.
- Fetterman, D. M., Delaney, L., Triana-Tremain, B., & Evans-Lee, M. (2015). Empowerment evaluation and evaluation capacity building in a 10-year tobacco prevention initiative. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Fetterman, D. M., & Eiler, M. (2001). *Empowerment evaluation and organizational learning: A path toward mainstreaming evaluation*. American Evaluation Association.
- Fetterman, D. M., Kaftarian, S., & Wandersman, A. (2015). *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Fetterman, D. M., & Wandersman, A. (Eds.). (2005). *Empowerment evaluation principles in practice*. Guilford Press.
- Ghoshal, S. (2005). Bad management theories are destroying good management practices. *Academy of Management Learning & Education*, 4, 75–91.
- Greenwood, D. J., & Levin, M. (1998). *Introduction to action research: Social research for social change*. SAGE.
- Labin, S., Duffy, J. I., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2013). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33(3), 307–338.
- Langhout, R. D., & Fernandez, J. S. (2015). Empowerment evaluation conducted by fourth- and fifth-grade students. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Lewin, K. (1951). Problems of research in social psychology. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers*. Harper & Row.
- Mithaug, D. E. (1991). *Self-determined kids: Raising satisfied and successful children*. Macmillan.
- Mithaug, D. E. (1993). *Self-regulation theory: How optimal adjustment maximizes gain*. Praeger.
- Patton, M. Q. (1997a). Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice*, 15(3), 311–320. Available at www.davidfetterman.com/pattonbkreview1997.pdf.
- Patton, M. Q. (1997b). *Utilization-focused evaluation: The new century text* (3rd ed.). SAGE.
- Patton, M. Q. (2013). The roots of utilization-focused evaluation. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (pp. 293–303). SAGE.
- Patton, M. (2015). *Review of empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability* (2nd ed.). (Evaluation Program and Planning, X). SAGE.
- Ployhart, R. E., & Bartunek, J. M. (2019). Editors' Comments: There is nothing so theoretical as good practice—a call for phenomenal theory. *Academy of Management Review*, 44(3), 493–497.

- Preskill, H., & Torres, R. T. (1999). *Evaluative Inquiry for learning in organizations*. SAGE.
- Sastre-Merino, S., Vidueira, P., Diaz-Puente, J. M., & Fernande-Moral, M. J. (2015). Capacity building through empowerment evaluation: An Aymara women artisans organization in Puno, Peru. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. SAGE.
- Senge, P. (1994). *The fifth discipline: The art and practice of the learning organization*. Doubleday.
- Wandersman, A., Snell-Johns, J., Lentz, B., Fetterman, D. M., Keener, D. C., Livet, M., . . . Flaspohler, P. (2005). The principles of empowerment evaluation. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice*. Guilford Press.
- Zimmerman, M. A. (2000). Empowerment theory: Psychological, organizational, and community levels of analysis. In J. Rappaport & E. Seldman (Eds.), *Handbook of community psychology*. Kluwer Academic/Plenum Press.
- Zimmerman, M. A., Israel, B. A., Schulz, A., & Checkoway, B. (1992). Further explorations in empowerment theory: An empirical analysis of psychological empowerment. *American Journal of Community Psychology*, 20(6), 707–727.
- Zimmerman, M. A., & Rappaport, J. (1988). Citizen participation, perceived control, and psychological empowerment. *American Journal of Community Psychology*, 16(5), 725–750.

CHAPTER 18

The Curious Case of Practical Participatory Evaluation Explorations of the Theory–Practice Relationship

J. Bradley Cousins
Hind Al Hudib

INTRODUCTION

We congratulate Marvin Alkin and Christina Christie for producing this third edition of the amazing *Evaluation Roots* compendium. To our knowledge, no one has provided a more comprehensive or detailed mapping of the evaluation theory landscape. That this body of work represents an extraordinary and unparalleled contribution to our field would be a woeful understatement. The first two editions traced evaluation theorists' views and perspectives using the evaluation theory tree as an organizing device. In the second edition, careful consideration was given to a wider, more global perspective. We are delighted that in the current edition, the focus is on the relationship between theory and practice, and that we have been invited to share our own thoughts about this connection with regard to our own research, or dare we say, theory.

In both prior editions of *Evaluation Roots*, the editors located one of us (Cousins) on

the “use” branch of the evaluation theory tree, and with good reason. Our work on practical participatory evaluation (PPE), with its focus on program problem solving and improvement, was very much linked to overarching concerns about enhancing the usefulness of evaluation. In PPE, trained evaluators work in partnership with members of the program community¹ to co-produce evaluative knowledge. That knowledge has to do with the extent to which programs are effective and, more importantly, which aspects of the program contribute to its effectiveness. Such could be used to identify and implement program improvements (instrumental use), to deepen program community members' understanding about how programs operate and the contexts within which they operate (conceptual use), or, less importantly, to satisfy accountability demands for program review (symbolic use). Along the way, program community members' participation in conducting the evaluation could help in

their own development of evaluation knowledge and skills (process use).

As described in Cousins's (2004) initial chapter in the *Roots* collection, our theoretical contributions on PPE grew out of the respective and independent evaluation practice experiences of Cousins and Lorna Earl in the early 1990s. We came to the realization that our evaluation projects were not only being noticed by educational program stakeholders but also results were actually being used. We decided to write about this and published an initial theoretical article-length foray into PPE (Cousins & Earl, 1992). Not long afterward, we published perhaps the most widely cited text on PPE (Cousins & Earl, 1995), which included a theoretical, front-end chapter, a collection of empirical studies, and a concluding review and integration of the preceding studies. Another piece, which elaborated the relationship between participatory practices to evaluation use, appeared in an international handbook (Cousins, 2003). In each of these contributions, we argued that literature on evaluation utilization and organizational learning provides theoretical support for PPE stemming primarily from the view that knowledge is socially constructed and cognitive systems and memories are developed and shared by organization members. PPE was framed as an organizational learning system. Additionally, we observed that a body of empirical research supporting these theoretical postulates was accruing.

Despite the emergence and recognition of PPE and its theoretical underpinnings (e.g., Alkin & Christie, 2004; Hansen, Alkin, & Wallace, 2013), the evolution of our ongoing thinking about participatory evaluation was dramatically influenced by a chapter we produced called "Framing Participatory Evaluation" (Cousins & Whitmore, 1998), which captured the interest of many evaluation scholars and practitioners at least partly because it explicated and clarified the existence of two streams of inquiry in participatory evaluation, each having different roots,

intentions, and justifications (King, 2007). Cousins and Whitmore (1998) succeeded in revealing the enormous range of evaluation practice that fell under the term *participatory*. Much of this work had roots in international development, was ideological in form and function, and its ties to a social justice agenda connected only tangentially with evaluation utilization and organizational learning considerations. Hence the identification of separate but related streams of PPE and transformative participatory evaluation (TPE). Perhaps the closest representation of the latter stream on the evaluation theory tree is Mertens's (2009) transformative research and evaluation model. Mertens is located on the "valuing" branch of the theory tree, quite distant from PPE's location on the "use" branch (Christie & Alkin, 2013). It is interesting that these two approaches should be located on different branches despite a common defining feature: trained evaluators working in tandem with program community members to produce evaluative knowledge.

Over time, our research program has evolved to focus on an overarching domain of inquiry called *collaborative approaches to evaluation* (CAE), of which both PPE and TPE are member cases. CAE is an umbrella term that encompasses a wide range of collaborative approaches that share the basic evaluator–stakeholder partnership feature. These approaches are quite diverse in intent and form. Over time, we developed and validated a set of evidence-based principles to guide CAE (Cousins, 2020; Cousins, Whitmore, & Shulha, 2013; Shulha, Whitmore, Cousins, Gilbert, & Al Hudib, 2016). These principles encapsulate the essence of our current "evaluation theory" and provide an overarching frame for this chapter. Perhaps it is fair to say that work on CAE is a natural evolution of our contributions to thinking about PPE. Yet, realistically, this evolution was brought about by situating PPE in a larger participatory context: a merger of use-oriented and social justice values-oriented

participatory perspectives. In this chapter, we consider how best to enable evaluation practitioners’ and indeed stakeholders’ use and application of the CAE principles in practical program evaluation contexts.

Yet, for the purposes of this compendium, PPE is a curious case and perhaps no longer a good fit for the theory tree. Just as a tree is a living, growing organism, we might similarly think of evaluation theories. Theories can evolve and change over time, and this is most certainly the case with PPE. As we explain below, we eschew efforts to differentiate among members of the CAE family of approaches (that includes PPE) in favor of investing heavily in an analysis of context to determine (1) whether a collaborative approach makes sense and, if so, (2) what it should look like in practice. In essence, we are now disinclined to promote specific collaborative approaches in and of themselves. So, while PPE may be locatable on the use branch of the theory tree, its underlying theoretical tenets arise from evidence-based CAE principles, a theoretical framework that defies positioning on the theory tree by virtue of the enormous diversity among family members.

Before we explicate the development, specification, and intended consequences of the CAE principles, we share some views about the theory–practice relationship in the social sciences by way of situating our own work. The chapter culminates in some ideas about how to enable practical use and application of the CAE principles and some thoughts about how PPE fits into all of this.

THEORY–PRACTICE RELATIONSHIP

Theory into Practice or Vice Versa?

Many years ago, one of us (Cousins) attended the universitywide lecture on reflective practice delivered at the University of Toronto by the eminent Massachusetts Institute of Technology (MIT) Professor Donald Schön. In his introduction to the invited

speaker, Professor David Hunt reminded us of psychologist Kurt Lewin’s (also formerly at MIT) famous adage: “There is nothing as practical as a good theory” (1945, p. 129). Hunt raised this observation because his impression of Schön’s work on the “reflective practitioner” suggested very much the opposite: *There is nothing as theoretical as good practice*. I found this observation to be really quite prophetic; to that point, my understanding of theory building was very much defined by the stepped process of (1) deep observation of real-life practice, (2) inductive theory building from practice cases, and (3) deductive testing of inductive theory using more practice cases. To the extent that research shows the theory to be valid, it can be used to inform practice and research. To the extent that the theory is good, it is practical. I soon came to realize that a movement in how we think about theory building was afoot.

Hunt (1987) later went on to write about this inverted perspective and over the years, and even quite recently, others have embraced it in very direct ways. For example, Huarng and Mas-Tur (2016) published a special peer-reviewed journal issue “Turning Kurt Lewin on His Head,” drawing from 75 papers at an international conference on global innovation and knowledge. In management studies, the associate editors of the *Academy of Management Review* used the inverted maxim as the basis for a call for phenomenological theory (Ployhart & Bartunek, 2019). They lamented how impact factors reveal that peer-reviewed journals that publish theoretical papers are doing quite well, but that “if good theory is also practical, then the record of success is far less certain” (p. 493).

In prior editions of *Evaluation Roots*, I have championed the role of systematic empirical inquiry in evaluation as the bridge between theory and practice (Cousins, 2004, 2013a). That bridge can be used for different purposes. From a traditional perspective, it could be used deductively to test inductive

theories or theoretical propositions. From the foregoing alternative perspective, it could be used as a means to harvest or make sense of good practice, with the end goal of developing theory and deepening understanding of complex evaluation practices. It is within this latter perspective that our CAE research program resides.

When Does a Conceptual Framework Become a Theory?

In an invited paper in 2013, I raised the question: “When does a conceptual framework become a theory?” (Cousins, 2013b). I wrote the paper in response to a request by Marvin Alkin and associates, who at the time were compiling a special issue on visualizing evaluation theory. Their general approach was to take specific theories and to provide visual representations of them in the form of logic models. Their purpose in doing so was to make theories more accessible to students and evaluation practitioners. In the paper, I commented on the relatively fine job they did in representing PPE in this way, although I had a few relatively minor quibbles to share with them. However, my main concern was that PPE, in my mind, is not a theory, nor is it a model or prototype. Rather, we prefer to think of it as an approach, one that is free to take on varying purposes and goals, and to manifest itself in different ways, depending on the conditions and exigencies associated with the context within which the program resides. To that end, I subtitled the article “Reflections from an Accidental Theorist.” I consider myself a researcher, after all.

It is true; in our own research, we use visual representation for guidance. However, in doing so, we generate conceptual frameworks—collections of variables or constructs and suggested relationships among them—that help to bound inquiry in terms of instrument development, data collection, data analysis, and interpretations. Almost invariably, our research ends with implications for research and theory, *and* implica-

tions for practice; the consequences of our research ultimately move in both directions on the bridge as our findings deepen our understanding of complex phenomena.

Mark and Henry (2004) observed that our PPE conceptual framework—intended to guide inquiry about participatory evaluation antecedents, processes, and consequences—may have been more than just a collection of concepts and suggested relations among them. In their words, “Cousins (2003) makes the applicability of logic models to evaluation use more explicit than usual, offering a model of utilization that looks strikingly like a program logic model” (p. 37). Using the logic model analogy, they went on to adapt parts of the framework for their own model of use and influence. Therefore, I had wondered, “Had I done something similar? Had I implicitly formulated a logic model of PPE? Had I laid out a theory of PPE? If I had, I had done so unknowingly” (Cousins, 2013b, p. 70). We have since updated the framework and adapted it to a much broader array of participatory and collaborative approaches to evaluation (Cousins & Chouinard, 2012). Elaborated in this most recent version were considerations of context, which would accommodate a broader array of approaches, including those with transformative intent. Again, our interest was in guiding inquiry into participatory and collaborative approaches as opposed to offering some sort of prescriptive theory. Yet the question remains, when does a conceptual framework become a theory? At this writing, no clear answer is apparent to us.

In the ensuing section, we discuss the rationale for, and development and validation of, a set of principles to guide practice in CAE. In our minds, this set of principles, which includes recommendations for action by evaluators and others, is more clearly aligned with the requirements of prescriptive theory. What is more, this theoretical framework emerged from a systematic examination of practice; that is to say, it is not only evidence-based but it also is consistent with a practice-into-theory orientation.

THE ESSENCE OF CAE

Rationale, Goals, and Processes

Many current evaluation approaches can be classified as collaborative by virtue of knowledge production being a function of the joint work of evaluators and program community members. In addition to PPE, mentioned earlier, we could add contribution analysis, participatory action research, empowerment evaluation, model of transformative research and evaluation, democratic evaluation, most significant change technique, and utilization-focused evaluation (U-FE), among a host of other approaches. Note that some of these approaches are by definition collaborative (e.g., empowerment evaluation, transformative participatory evaluation [TPE]), whereas others (e.g., U-FE, collaborative evaluation) are more or less collaborative depending on evaluation context and implementation choices.

While some contributors (e.g., Fetterman, Rodríguez-Campos, Zukowski, & Contributors, 2017) choose to focus on differentiating one collaborative approach from another, we see this objective as being wrongheaded. With such a rapid proliferation of evaluation approaches, all sharing the common feature of evaluators working in tandem with program community members to conduct evaluation, we saw good value in the prospect of an overarching set of principles to guide practice. Another critical part of this argument is that the goals and objectives of any given CAE and the form that it takes in practice ought to arise from a thorough understanding of program context. We eschew the notion of evaluators self-identifying with one particular CAE and looking for practice opportunities where the approach would fit (Cousins et al., 2013; Cousins, Shulha, Whitmore, Al Hudib, & Gilbert, 2020).

The overarching goal of our evidence-based principles to guide CAE (described below) is to enable practicing evaluators and their program community-based colleagues to decide whether CAE should be employed in the first instance, and if so,

to help them to implement the evaluation in the most effective and suitable way possible. This includes identifying goals and purposes, deciding appropriate processes, invoking quality control safeguards, and enhancing impact.

It is important to consider when to choose CAE in response to the presenting evaluation challenges, and three possibilities come to mind (Cousins, 2020). First, CAE is appropriate to use when the impetus for program evaluation is formative or summary formative. This is not to say that CAE should not be used to assess program impact, but realistically in situations where “fork in the road” judgmental questions are being asked (e.g., fund program or terminate, expand program or not), its application is not advisable. In any CAE, part of the evaluator’s role is to ensure that professional standards of evaluation practice are upheld. In situations where program funding or community members’ livelihoods are on the line, it may not be realistic to expect the evaluator to monitor and control stakeholder self-serving bias. A second opportunity to use CAE arises when information needs are developmental. In developmental evaluation (Patton, 2011) the evaluator is at the decision-making table with an array of stakeholders, where the nature and scope of innovative approaches to social, educational, or health problems are being considered. The evaluator’s role is to provide evidence-based support to ongoing deliberations. Developmental evaluation, by definition, has a natural fit with CAE. Third, CAE is entirely suitable to evaluations that have transformational intent (e.g., Mertens, 2009). In such cases, a significant purpose of the evaluation is to foster among intended program beneficiaries, often from marginalized populations, greater awareness of the complexities of their circumstances; knowledge and skills building; the restoration and preservation of culture; improved social, health, and economic standing; sustainability; and self-determination. It is perhaps more through the direct engage-

ment of such persons in the evaluation of interventions pertaining to them, than the actual evaluation findings, that these objectives can be achieved. Culturally responsive approaches can take on this aura: “Culturally responsive evaluation with indigenous peoples moves evaluation away from a collision course as evaluators and indigenous peoples become collaborators on evaluations, generating inquiry paradigms that draw upon indigenous values, knowledge, experiences and understandings” (Chouinard & Cram, 2020, p. 57).

Historically, we have identified three principal rationales or justifications for CAE (Cousins & Whitmore, 1998). We call these the three P’s. The *pragmatic* justification for CAE is salient when the information needs are about program management, improvement, and/or problem solving. We can think of the instrumental and conceptual use of findings as intended consequences. The *political* justification comes into play when there is significant emphasis on the amelioration of social inequity or the empowerment of marginalized groups. In such cases, CAE may be much more about the process as opposed to the results of the evaluation. Finally, the *philosophical* rationale associates with our need to understand complex program phenomena at deeper levels. This justification is grounded in epistemological considerations such as insider–outsider perspectives. We have argued that, depending on circumstances, the three P’s will be differentially emphasized in any CAE. Whereas PPE is predominantly justified by pragmatic considerations, TPE would be supported by political arguments. In both cases, partial justification would come from the other two rationales.

On reflection, implicitly, each of these rationales touches on ethical considerations. Recently, the import of such considerations has become apparent to us; thus, we have expanded our thinking about why to engage with CAE so as to include an “ethic of engagement” (Chouinard & Cousins, 2021). This we unfold along a set of interconnected

dimensions that we call the seven R’s: *reflexivity*, *relationality*, *responsibility*, *recognition*, *representation*, *reciprocity*, and *rights* (see Box 18.1). While the ethical justification for collaboration dynamically overlaps with the three P’s rationales, and perhaps especially with the political justification, what differentiates it is its focus on the stance or role of the evaluator in acknowledging diverse others and in creating a favorable space for collaboration to flourish. The seven R’s may be backgrounded in cases where the emphasis is on pragmatic, political, and/or philosophical reasons for doing CAE, or they may be foregrounded when the motivation for collaboration is driven by policies and programs imbued with ethical concerns. An example of the latter might be the evaluation of interventions intended to reconcile

BOX 18.1. Seven R’s of Collaborative Practice

- **Reflexivity** situates the evaluator as an active, conscious being in the evaluation process—self as responsive instrument.
- **Relationality** brings focus to the interdependent connection evaluators have with participants, as they focus on building a learning community together.
- **Responsibility** speaks directly to the purpose of human inquiry and to our obligations to others in ameliorating historical and ongoing injustices and inequities.
- **Recognition** refers to the inclusion of diverse others and to the creation of a “difference-friendly” society.
- **Representation** is about questions of voice and perspective, and who has a right to speak for others, and reciprocity brings attention to the relational and dialogic process of inclusion.
- **Reciprocity** emphasizes the relational, dialogic, and political aspects of collaborative practice—the give-and-take aspects. Who benefits from the partnership? Who does not?
- **Rights** refers to the protection of legal, civil, moral, or conventional obligations.

historical colonial injustices to indigenous peoples.

The form that CAE takes in practice will be a function of the combination of three fundamental dimensions: *control* of technical decision making (evaluator controlled—balanced—stakeholder controlled); stakeholder *diversity* (highly diverse group participating—midrange—only primary users participating); and *depth* of participation in technical evaluation activities (deep involvement midrange—consultative involvement) (Cousins & Whitmore, 1998). CAE is a dynamic process that may change through the course of a particular project. For example, the inclusion of a wide array of program community members may evolve into the active participation of only a select group as the evaluation unfolds. Given the complexities of working with large groups, this eventuality may not be entirely unexpected. Yet this does not imply that CAE needs to be restricted to small-scale evaluations. We recently completed a nationwide participatory evaluation of teacher inservice training in India involving over 50 educators and professors from nine states across the country (Cousins, Al Hudib, & the RMSA Team, 2015). Members faithfully participated in planning, instrument development, data collection, data analysis and interpretation, and reporting activities as the evaluation progressed.

Regardless of who is leading the evaluation—be they evaluators, program community members, or some mix of the two—CAE benefits from the involvement of partners from different roles and perspectives. The evaluator’s role is predominantly tied to their expertise in evaluation logic and methods and professional standards of practice, although they may have some knowledge of the intricacies of the program and the context within which it is operating. On the other hand, stakeholders participating in CAE may have some limited knowledge of evaluation logic and professional standards of practice, but their strong suit will be in understanding the program logic,

its history, and the context within which it is operating. The relative contribution from these different perspectives helps to define the power of CAE to accomplish its objectives. Regardless, we would argue that all of these practice-based actors stand to benefit from a thorough understanding of evidence-based principles for CAE. We now turn to a description of the principles, how they were developed, and how they may be used. Readers will find further detail in Shulha and colleagues (2016) and Cousins and colleagues (2020).

Figure 18.1 is our visual representation of the principles and Box 18.2 is an abbreviated description of each principle. Four observations about the principles are noteworthy. First, the eight principles in Figure 18.1 are to be used as a set as opposed to a pick-and-choose menu. Second, the dotted lines indicate that there is overlap among the principles. While each principle is relatively distinct, they are interconnected and not mutually exclusive. Third, while there exists a loose temporal ordering of the principles, beginning with “clarify motivation for collaboration,” they are not intended to be used in a stepwise, linear fashion. Rather, application of the principles should be iterative. Finally, some of the principles may also apply to conventional or mainstream approaches to evaluation. However, in this case, they were identified as being of central importance to the conduct of CAE and are therefore highlighted here.

Another characteristic of the set is that it is evidence-based. Consistent with our embrace of the practice-into-theory orientation toward theory building (described earlier), our approach to developing the principles was systematic and inductive. We conducted a multiphase study comprising two pilot phases, one main data collection phase, and a follow-up validation phase (see Shulha et al., 2016). In the main phase, we obtained quantitative and qualitative data from 320 practicing evaluators who are members of one or more of three professional evaluation societies.²

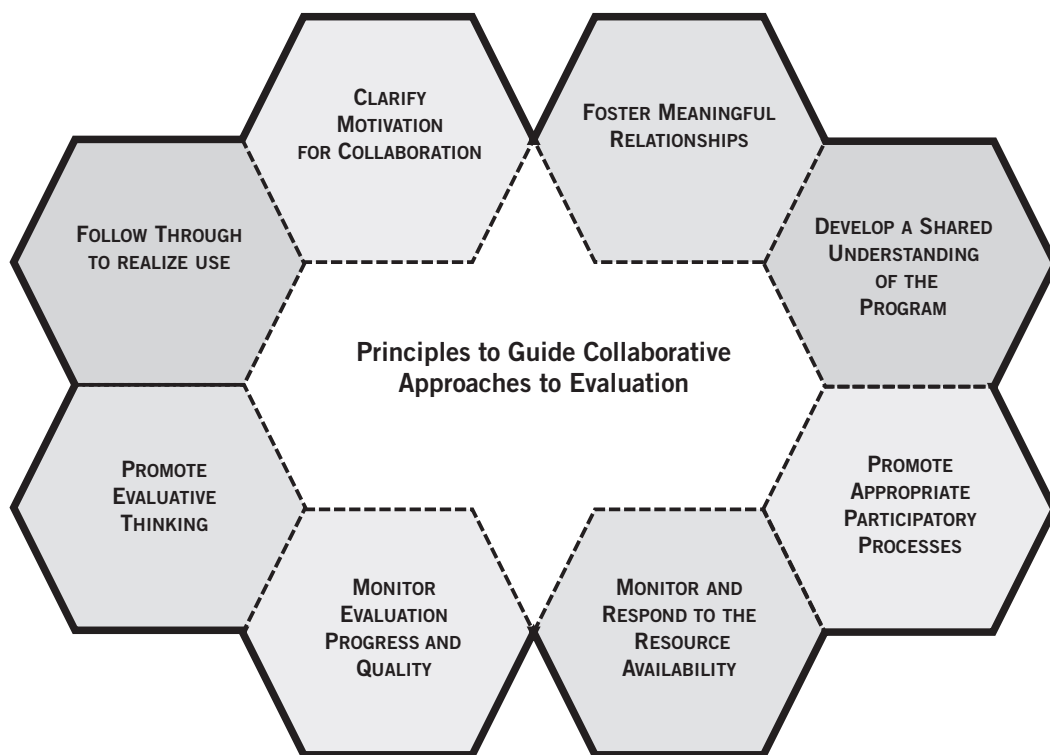


FIGURE 18.1 An integrated set of evidence-based principles for use in guiding collaborative approaches to evaluation. From Shulha et al. (2016). Copyright © 2016, SAGE Publications. Reprinted by permission.

We had each evaluator identify two specific CAE projects on which they had worked within the past 3 years. One of these was selected as being “highly successful” criteria for success being left to the respondent. The other was selected as being “far less successful than hoped,” again according to respondent expectations. Qualitative data associated with each project (counterbalanced for order) were gathered in response to the question “What were the three most important reasons why this project was [successful/less successful than hoped]?” (counterbalanced for order). From these data, we identified “themes,” higher-order “contributing factors”, and ultimately the eight CAE principles. We then circulated materials to a volunteer subsample of 51 respondents and had them complete a validation exercise (fourth phase of the study). This process enabled the emergence of the principles

from the successful and less-than-successful experiences of practicing evaluators.

We consider this version of the principles to be a solid preliminary formulation. However, we have continued to subject the set to empirical field tests, with an eye toward revising them based on what we learn through systematic inquiry (Cousins, 2020).

We envisioned five specific applications of the principles, although this list is not exhaustive. First and foremost, the principles could provide guidance to practicing evaluators and program community members engaged in planning, designing, and implementing CAE projects. Second, the principles provide a framework for the retrospective review of already completed CAE projects. Third, the principles could serve as the basis for instruction in either graduate study or practical evaluation capacity-building workshops or short courses.

Fourth, they could be used by organizations to develop or revise existing evaluation policies. Finally, the CAE principles could provide a conceptual framework to guide ongoing inquiry into the antecedents, practices, and consequences of CAE.

BOX 18.2. Brief Description of CAE Principles

Clarify motivation for collaboration: Why use CAE? The principle encourages the development of a thorough understanding of the justification for the collaborative approach.

Foster meaningful relationships: The principle inspires the conscious development of quality working relationships between evaluators and program stakeholders and among stakeholders, including open and frequent communication.

Develop a shared understanding of the program: Is the program commonly understood? Is everyone in agreement about the intended program processes and outcomes? The principle promotes the explication of the program logic situated within context.

Promote appropriate participatory processes: The principle encourages deliberate reflection on the form that the collaborative process will take in practice with regard to specific roles, and responsibilities for the range of stakeholders identified for participation.

Monitor and respond to resource availability: The principle underscores the critical importance of data quality assurance and the maintenance of professional standards of evaluation practice.

Promote evaluative thinking: The principle inspires the active and conscious development of an organizational cultural appreciation for evaluation and its power to leverage social change.

Follow through to realize use: To what extent is evaluation a valuable learning experience for the stakeholder participants? The principle promotes the conscious consideration of the potential for learning, capacity building, and other practical and transformative consequences of the evaluation. Implicated are evaluation processes and findings.

CAE IN PRACTICE

Once we had developed and validated the principles, we considered a strategy for launching them within evaluation communities around the globe. We had the principles translated into French and Spanish, and worked with colleagues where those languages are spoken to raise awareness and encourage dissemination. We officially launched the principles in January 2017 through various evaluation society news outlets and the listservs. The launch included two appended documents: (1) a brochure-style document that provided a brief description of the principles, their intended uses, and contact information and (2) an indicators document that provided readers with suggested actions and questions to ask associated with each of the principles. We also put together a short video (Cousins et al., 2016) in which our research team provided a brief overview of the principles and their intended applications. Finally, leading up to the launch, and subsequent to it, we engaged in 19 significant promotional/training activities such as workshops and conference presentations in North America, Europe, Middle East, and South America. Throughout our travels, we encouraged participants to consider developing ideas and proposals for field-testing the principles. These efforts culminated in the production of a compendium of eight peer-reviewed field studies (Cousins, 2020).

While the foregoing speaks to our ongoing efforts to promote the principles, of course, the challenge comes in fostering evaluator and stakeholder understanding and application of them in practice. Of high interest would be the use of the principles to guide CAE practice from the outset. To that end, we developed the indicators document mentioned earlier. That document comprises a brief description of the principles alongside actions evaluators can take and questions that can be pondered. These actions and questions (indicators) appear in Table 18.1.

TABLE 18.1. Actions and Indicators to Support CAE Principle Application

Principle	Action	Indicators
Clarify motivation for collaboration	Assess through consultation, documentation, examination who are the important stakeholders and what they value	• Is there a general high level of agreement among stakeholders about program information needs? • Are program goals consistently embraced across stakeholders? • Are identified evaluation purposes weighted toward formative improvement orientation? Are there developmental or transformative objectives?
Foster meaningful relationships	Openly discuss, listen, and negotiate decision-making processes and boundaries	• What should be the composition of stakeholder participation? • Are regularly scheduled meetings held; are they having consistent attendance; are there respectful and professional deliberations? • Is the process inclusive, with all members' voices included?
Develop a shared understanding of the program	Initiate exercise to develop and validate program logic model	• Is the evaluator helping stakeholders teach them the program's logic? • Can reasonable consensus among stakeholders about the program logic be reached? • What is the program's status within the organizational context?
Promote appropriate participatory processes	Negotiate what form the collaboration will take	• Who will control evaluation technical decision making? • To what extent are stakeholders involved in all evaluation activities? Are they comfortable with their level of participation? • Is the involvement of stakeholders sufficiently diverse? Are distinct perspectives represented?
Monitor and respond to resource availability	Negotiate resources for the evaluation in advance; periodically reevaluate during process	• Is there a memo of understanding or contract? • Is there alignment between evaluation scope and required funds, time, and personnel? • Can this be renegotiated in the face of unanticipated developments?
Monitor evaluation progress and quality	Take steps to ensure data quality as per professional standards of practice	• Do stakeholders understand the critical importance of data quality assurance? Is some level of instruction/demonstration required? • Are there competing demands marshaling against identified standards/procedures? Can they be negotiated?
Promote evaluative thinking	Assess facility and comfort with systematic inquiry and develop informal plan for growth and development	• Has the evaluator used teachable moments, critical events to convey the value of inquiry-oriented culture? • Are stakeholders aware of and appreciative of learning that has taken place (findings and process)? • Do stakeholders see value in systematic inquiry for other aspects of their role? • Have stakeholders had ample opportunity to determine the implications of the findings?
Follow through to realize use	Plan to formally or informally follow through after the findings have been established	• Is there an opportunity for knowledge mobilization within the organization? • Have stakeholders' roles and responsibilities changed in significant ways? What supports are needed to help with such changes? • Have all implications and recommendations arising from the evaluation been given fair treatment?

Note. From Cousins (2020). Copyright © 2020, SAGE Publications. Reprinted by permission.

In order to enhance the prospect of effective CAE practice, it is paramount that evaluator and stakeholder participants take the time to review and deliberate these actions and questions. Critical among these are issues associated with “Clarify motivation for collaboration.” Unless the choice to use CAE can be justified from pragmatic, political, philosophical, or ethical perspectives, practitioners ought to consider alternative, noncollaborative approaches. The principle “Foster meaningful relationships” is all about developing sustainable and effective working relations. The potential for conflict will inevitably be there; evaluators need to understand and appreciate the diversity among those program community members participating in the process. Through the specified action associated with “Develop a shared understanding of the program,” the evaluator will rely on the practical insight and wisdom of participating stakeholders. However, they should seek to augment this with their own understanding of generic program structure and function. Decisions around “Promote appropriate participatory processes” will define the shape the collaborative process will take. Who will be making these decisions and why? Evaluation and program practitioners ought to keep in mind the three dimensions of form elaborated earlier: control, diversity, depth of participation. What would be the appropriate balance on each of these spectra and why?

Both evaluator and program community members have roles to play in “Monitor and respond to the resource availability.” Evaluators should be well positioned to estimate costs for specific evaluative activities, whereas program practitioners will have a clearer idea about resource availability and constraints. Stakeholders ought to defer to evaluators’ expertise with regard to “Monitor evaluation progress and quality” given that the latter ought to be well versed in professional evaluation logic and standards of practice. Program practitioners can help to situate data quality-assurance strategies

within the local context. Opportunities to “Promote evaluative thinking” and nurturing process use throughout the evaluation are abundant. Evaluators ought to take whichever opportunities present themselves to foster the learning and development of participating stakeholders and to encourage the development of an organizational culture of inquiry and learning. Finally, post-evaluation communication will be important to ensure the intended impact of the CAE project. Both evaluators and stakeholders should “Follow through to realize use.” It is important to recognize that *use* in this context may imply use of findings, processes, transformational benefits, and/or development advances, depending on the goals and purposes of the particular CAE.

THE CURIOUS CASE OF PPE

Through the lens of our CAE theoretical perspective, as would be the case for any collaborative approach, the purposes, methods, implementation, and follow-up of PPE would arise from a thorough analysis of the context for the evaluation. We cannot emphasize enough the application of the “Clarify motivation for collaboration” principle in making such determinations. PPE would be likely to take shape where answers to the indicator questions point toward a situation where program information needs are agreed, and they align with formative, improvement-oriented interests. Once that determination has been established, subsequent principles would be applied in a manner consistent with the problem-solving, use-oriented purposes of PPE. Hypothetically, stakeholder participation might be limited to primary users—stakeholders with a vested interest in the program and its evaluation—who can actually act on emergent recommendations for program change. Decision making about technical aspects of the evaluation would be balanced between evaluators and stakeholders, and stakehold-

ers would be likely to engage in a full range of evaluation activities, including data processing, analysis, and interpretation, as well as reporting. Organizational commitment to the evaluation would be required to ensure that participating stakeholders are free to engage in the inquiry. The evaluator would ensure that professional standards of evaluation practice are respected, and resist hints of stakeholder self-serving bias. By virtue of process use through direct participation, it is likely that evaluative thinking would be promoted within the program community. Integration of evaluation into the organizational culture might be enhanced in cases where the program community experiences successful use of the evaluation. Finally, evaluators and stakeholders might work together to ensure that findings are understood by key decision makers and recommended actions given due consideration. In that primary users are directly engaged with the evaluation, follow-up would be more natural than would be the case with an externally conducted evaluation; ownership of the findings and the process would be evident among participating stakeholders.

The foregoing is merely a hypothetical portrait of what the evaluation might look like if the impetus for evaluation called for PPE. It is important to remember that the application of the principles is iterative and that circumstances may change throughout the course of the evaluation. But what is critical to understand is that the decision to implement a PPE evaluation needs to be justified on the basis of a nuanced reading of context, and that that reading will ultimately inform decisions about what the collaborative approach will look like and what effects it may be expected to have.

FINAL THOUGHTS

We conclude with three final thoughts about the role of empiricism in moving CAE theory into practice: the value of a research agenda, research as an observational record,

and the reciprocity of the theory–practice relationship. First, ongoing inquiry should be informed by an agenda for research, one that addresses important gaps in research-based knowledge. Two important gaps spring to mind, although many others exist. The compendium of field studies we have intentionally gathered so far (Cousins, 2020), and others that have been published independently (e.g., Searle, Merchant, Chalas, & Lim, 2017; Shulha, Searle, Poth, & Chalas, 2019) have focused predominantly on retrospective analysis and pedagogical applications with only limited attention to the use of the principles as guidance for practice. A study by Acree (2019) is a good example of the latter. He used the CAE principles the planning, implementation, and management of a collaborative evaluation of a technology integration program intended to enhance K–12 teacher preparation. Such applications should be of high interest in the recruitment and development of ongoing field studies.

Many of the CAE field studies conducted to date have been exclusively from the perspective of the evaluator. Yet CAE, by definition, implicates the role, involvement, and perspectives of participating stakeholders. Shulha and colleagues (2019) explored stakeholder perceptions of CAE through a survey of 25 program managers or implementers who had previously participated on such projects. The study provided rich detail about stakeholders' experiences with and views about roles and responsibilities, learning, and the development of professional relationships. It also provided insights into how to access members of the program community for inquiry. This study serves as an excellent model for broadening perspectives in ongoing inquiry into CAE practice.

Our second point is that empirical studies of CAE practice, particularly those framed by CAE principles, contribute to an accumulating record of practice observation and such a record can be a rich source of ideas and inspiration for practicing evaluators and stakeholders. However, most such

research is published in scholarly outlets, such as peer-reviewed journals or edited books, and may be relatively inaccessible to practitioners. We would suggest that serious consideration be given to knowledge mobilization strategies that aim to translate and disseminate research-based knowledge about CAE theory and practice in ways that are more accessible to those involved in practice. Such strategies might include the use of knowledge brokers and the publication of research digests or articles in practice-based magazines and outlets. Implicated may be a role for social media, particularly forms that embrace two-way communication potential such as blogs, chat rooms, discussion/support groups, and the like.

Our final idea is that there may be something reciprocal about theory building; what goes around comes around in the theory–practice relationship. While the initial development of the CAE principles was inductive and grounded in the experiences of practicing evaluators (i.e., practice into theory), CAE theory will continue to evolve and refine on the bases of systematic inquiry into practice-based applications. It might be argued that marshaling inquiry along these lines ought to be construed as CAE theory building from a theory-into-practice orientation; and that is just fine by us. We delight in the prospect of traffic (knowledge, wisdom, insight) on the CAE bridge of inquiry continuing to move in both directions. We would argue that what is important is to value the power and potential of systematic inquiry to advance CAE theory and practice.

ACKNOWLEDGMENTS

Research on the development and validation of the CAE principles reported here was conducted by the members of the COVE research team: Brad Cousins, Bessa Whitmore, Lyn Shulha, Hind Al Hudib, and Nathalie Gilbert. Financial and in-kind support was provided by the University of Ottawa.

NOTES

1. In this chapter, we use the terms *program community member* and *stakeholder* interchangeably; they both imply social, educational, or health program funders, developers, managers, implementers, intended beneficiaries, special interest group personnel, and others.
2. Mostly the American Evaluation Association (AEA), secondarily from the Canadian Evaluation Society (CES), and the International Development Evaluation Association (IDEAS).

REFERENCES

- Acree, J. (2019). Applying the collaborative approaches to evaluation (CAE) principles in an educational evaluation: Reflections from the field. *Canadian Journal of Program Evaluation*, 34(2), 272–281.
- Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 12–65). SAGE.
- Chouinard, J. A., & Cousins, J. B. (2021). Developing an ethical rationale for collaborative approaches to evaluation. *Evaluation*, 27, 364–381.
- Chouinard, J., & Cram, F. (2020). *Culturally responsive approaches to evaluation: Empirical implications for theory and practice*. SAGE.
- Christie, C., & Alkin, M. C. (2013). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots* (2nd ed., pp. 11–57). SAGE.
- Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245–265). Kluwer.
- Cousins, J. B. (2004). Crossing the bridge: Toward understanding use through empirical inquiry. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 319–330). SAGE.
- Cousins, J. B. (2013a). Privileging empiricism in our profession: Understanding use through systematic inquiry. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (pp. 344–370). SAGE.
- Cousins, J. B. (2013b). When does a conceptual framework become a theory?: Reflections

- from an accidental theorist. *Evaluation and Program Planning*, 38, 67–70.
- Cousins, J. B. (Ed.). (2020). *Collaborative approaches to evaluation: Principles in use*. SAGE.
- Cousins, J. B., Al Hudib, H., & the RMSA Team. (2015). *Formative Evaluation of RMSA Teacher In-Service Training: Vol. 1. Main Report, Vol. 2. Case Studies, Vol. 3. Technical Documents*. RMSA Technical Cooperation Agency.
- Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: A review and integration of the research base*. Information Age.
- Cousins, J. B., & Earl, L. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(3), 397–418.
- Cousins, J. B., & Earl, L. (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. Falmer.
- Cousins, J. B., Shulha, L. M., Whitmore, E., Al Hudib, H., & Gilbert, N. (2020). Situating evidence-based principles to guide practice in collaborative approaches to evaluation. In J. B. Cousins (Ed.), *Collaborative approaches to evaluation: Principles in use* (pp. 3–45). SAGE.
- Cousins, J. B., Shulha, L. M., Whitmore, E., Gilbert, N., & Al Hudib, H. (2015). *Evidence-based principles to guide collaborative approaches to evaluation*. <https://crecs.uottawa.ca/publications/ten-minute-window/evidence-based-principles>
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation*. *New Directions in Evaluation*, 80, 3–23.
- Cousins, J. B., Whitmore, E., & Shulha, L. M. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7–22.
- Fetterman, D., Rodríguez-Campos, L., Zukowski, A., & Contributors. (Eds.). (2017). *Collaborative, participatory, and empowerment evaluation: Stakeholder involvement approaches*. Guilford Press.
- Hansen, M., Alkin, M. C., & Wallace, T. (2013). Depicting the logic of three evaluation theories. *Evaluation and Program Planning*, 38, 34–43.
- Huarnig, K.-H., & Mas-Tur, A. (2016). Turning Kurt Lewin on his head: Nothing is so theoretical as a good practice. *Journal of Business Research*, 69, 4725–4731.
- Hunt, D. (1987). *Beginning with ourselves: In practice, theory, and human affairs*. Brookline Books.
- King, J. A. (2007). Making sense of participatory evaluation. In S. Mathison (Ed.), *Enduring issues in evaluation: The 20th anniversary of the collaboration between NDE and AEA, New Directions in Evaluation*, 114, 83–86.
- Lewin, K. (1945). The research center for group dynamics at the Massachusetts Institute of Technology. *Sociometry*, 8(2), 126–136.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10, 35–57.
- Mertens, D. (2009). *Transformative research and evaluation*. Guilford Press.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.
- Ployhart, R. E., & Bartenuek, J. M. (2019). Editors' comments: There is nothing so theoretical as good practice—A call for phenomenal theory. *Academy of Management Review*, 44(3), 493–497.
- Searle, M., Merchant, S., Chalas, A., & Lim, C. Y. (2017). Collaborative approaches to evaluation in a development evaluation context. *Canadian Journal of Program Evaluation*, 31(3), 350–373.
- Shulha, L. M., Searle, M., Poth, C. N., & Chalas, A. (2019). Stakeholders weigh in on collaborative approaches to evaluation. In J. Chouinard, I. Bourgeois, & C. Amo (Eds.), *Growing the knowledge base: The contributions of J. Bradley Cousins* (pp. 99–122). Information Age.
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & Al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation*, 37(2), 193–215.

CHAPTER 19

Learning-Centered Evaluation

The Transformational Power of Evaluation Redux

Hallie Preskill

INTRODUCTION

It's all connected. This has been my mantra for a long time, and with each passing day, I believe it more strongly. One benefit of growing older is that you have the opportunity to reflect on why you do what you do and how you got where you are. For me, this has meant reflecting on my passion for evaluation—when it started, why I made certain educational and professional choices, and why I still think, after nearly 40 years in the field, that evaluation remains an exciting, challenging, and worthwhile enterprise.

THE ROOTS OF MY EVALUATION JOURNEY

It all began in a course on evaluation I was taking as part of a Certificate of Advanced Study (postmaster's) in 1979 at the University of Vermont. I had been a special education teacher for 4 years and was taking courses for no other reason than I liked learning new things. I had always enjoyed

doing research as part of my coursework and wanted to become a “researcher.” However, not having any mentors or particularly effective career guidance up to that point, I was unaware of how one actually got a job doing social science research. Though I enjoyed teaching, I knew there was something else out there for me, but I just didn't know what it was until I took a course on educational program evaluation, in which I read the work of Ralph Tyler, James Popham, Michael Scriven, Robert Stake, Barry MacDonald, Rob Walker, Ernie House, and Elliot Eisner. After only a few weeks, I knew my calling was evaluation. It was as if a whole new world had suddenly opened up to me.

By the end of the course, I had applied for and been accepted to do an internship at the University of Vermont's Center for Policy Research and Evaluation with Robert Carlson. I left my teaching job and gained invaluable experience working with veteran evaluators on several evaluations. During my year at the center, Robert Stake was invited to consult on an evaluation Carlson

was conducting. I had recently read Stake's 1978 article, "The Case Study Method in Social Inquiry," published in *Educational Researcher*. This piece, and others by Stake, profoundly influenced my thinking and future practice. Specifically, Stake taught me how case studies could be used to provide readers with vicarious experiences of program life, and how important it is to communicate the nuances of program participants' experiential understandings. I found Stake's discussion of tacit knowledge, naturalistic generalizations, and the notion that truth lies in the particulars both enlightening and inspiring. As I've reflected on my entrée to evaluation, I think I connected Stake's perspectives and theories of evaluation to what it means to learn. And *learning* has always been at the heart of my evaluation theory and practice.

Before leaving Vermont, Bob Stake offered me an opportunity to become a doctoral student at the University of Illinois (Urbana-Champaign), with an assistantship in the Center for Instructional Research and Curriculum Evaluation (CIRCE). That invitation and my subsequent studies and evaluation experiences at the University of Illinois constituted a major turning point in my professional life. Since finishing my graduate work in 1984, I have practiced evaluation in a wide variety of settings, as both an internal and external evaluator. After working in the corporate sector (as an evaluator) for a few years after my doctoral studies, I spent the next 22 years as an academic, teaching graduate-level courses in evaluation, research, training and development, organizational learning, and Appreciative Inquiry. Deciding that I wanted to become more steeped in practice again, I left the academic world in 2009, for the life of an evaluation and learning consultant.

For 12 years, I led (and co-led) the Strategic Learning and Evaluation practice at FSG, a nonprofit strategy, evaluation, and research organization, committed to supporting systems-level social change and equitable outcomes. While I loved being an

academic and all that it entailed, the last dozen years have been the most growth-oriented and challenging. During these years, I worked on over 150 evaluation and learning projects, with scores of foundation, nonprofit, and corporate philanthropy clients. As a result, my evolving theory of evaluation is firmly grounded in the observations and experiences that my colleagues and I have had over time. I have always considered myself a "theory-to-practice" professional—believing strongly in the significance of theory; however, my concerns with evaluation have always been practical. To this end, I've always strived to translate evaluation theory into practice in the hope of helping others value and engage in evaluative thinking and practice.

As I describe in this chapter, working as an evaluator committed to social change has required me to test my theory of evaluation every day. And it has helped me continue to evolve and shape my thinking and beliefs about the power and purpose of evaluation as a means for learning that ultimately influences transformative positive systems change.

AN EVALUATION THEORY THAT FOCUSES ON LEARNING AND EVALUATION USE

As a prelude to describing how my theory has evolved, and what I see as the underlying assumptions and practice guidelines of a learning-oriented evaluation theory, I offer the following parameters:

- The goal of my evaluation theory is that evaluation is a means for ongoing learning from evaluation processes and findings, and that, ultimately, evaluation contributes to decision making and action.
- Learning-oriented evaluation is probably best suited for developmental and formative evaluations, though one would hope that evaluators would facilitate some kind of learning experiences around summa-

tive evaluation processes and findings with stakeholders.

- Evaluation that intentionally builds in learning throughout the process is relevant to evaluands that are both small and large. However, stakeholder learning in multisite, complex, and geographies, is likely to be more complex and challenging—but doable.
- Learning-oriented evaluation is a generic theory—it is not specific to a particular topic, sector, geography, or type of program, organization, and community.

While my theory of evaluation has never veered far from its roots in use and learning, I do think it has evolved. As I reflect on my evaluation career, I can see this evolution in three phases, each one influenced by reading, writing, practice, and conversations with other evaluators and clients, observing where the evaluation field was in its own development, and sensing the dynamic political and social context of the environments in which we work.

The first phase reflects the work I did with Rosalie T. Torres and Mary Piontek, in which we explored the ways evaluation use could be enhanced through more creative and effective strategies for communicating and reporting evaluation findings, which resulted in the book *Evaluation Strategies for Communicating and Reporting* (1996). As we wrote this book, we were struck by how much of the communicating and reporting process is actually about learning. As Rosalie and I further explored the learning connection to evaluation, we developed a framework, and a subsequent book, *Evaluative Inquiry for Learning in Organizations* (1999), in which we described how social learning processes (asking questions, dialogue, identifying and exploring values, beliefs, and assumptions, and reflection) should be embedded in evaluation practice as a means for supporting individual, team or group, and organizational learning. Based on feedback we received from those who

read the book, we developed a diagnostic survey, *Readiness for Organizational Learning and Evaluation* (2000b; ROLE; www.fsg.org/tools-and-resources/readiness-organizational-learning-and-evaluation-instrument-role), to help organizations diagnose the strength and extent to which organizational learning and evaluation practices are present in their organizations. This learning-oriented evaluation approach has undoubtedly been the bedrock of my evolving evaluation theory and professional practice throughout my career.

The second phase of work that influenced my evolving evaluation theory is the work I did with Shanelle Boyle on evaluation capacity building (ECB) in 2007–2008. With the goal of helping others learn and appreciate evaluation, we undertook a research study in an effort to understand what was happening in the name of ECB, and to develop an integrated framework that could guide those who are designing and implementing capacity-building experiences for a range of audiences. This work culminated in a “multidisciplinary model” and the following definition:

ECB involves the design and implementation of teaching and learning strategies to help individuals, groups, and organizations, learn about what constitutes effective, useful, and professional evaluation practice. The ultimate goal of ECB is sustainable evaluation practice—where members continuously ask questions that matter, collect, analyze, and interpret data, and use evaluation findings for decision-making and action. For evaluation practice to be sustained, participants must be provided with leadership support, incentives, resources, and opportunities to transfer their learning about evaluation to their everyday work. Sustainable evaluation practice also requires the development of systems, processes, policies, and plans that help embed evaluation work into the way the organization accomplishes its mission and strategic goals. (Preskill & Boyle, 2008, p. 444)

It's been exciting to see how this topic has been studied over the last decade; other

models have emerged, and evaluators and practitioners are deepening their understanding of what it takes to build meaningful and sustainable evaluation capacity with varied participants and organizations. Continued work in this area—helping people learn and develop new ways of thinking and behaving, and what motivates them to engage in effective evaluation practice—is essential for the future of the field.

The third and most recent evolution of my thinking is reflective of my being more deeply embedded in practice. One way this has manifested is in what I believe is the relationship between strategy and evaluation. While I have always thought that evaluation should be *strategic* and should inform strategy, working closely with strategy consultants has helped me to reflect on, clarify, and articulate just what this means.

Not very different from other kinds of evaluation, the focus of strategic evaluation is on questions of critical importance to the organization's mission and goals, and positions evaluation as a means for collecting relevant, credible, and useful evaluation information for strategic decision making and action. However, strategic evaluation is primarily concerned with a set of questions that often differs from many program- or grant-level evaluations:

- To what extent are we making the right strategic choices?
- What are we learning about how well our programs are making progress on implementing our strategy?
- What else should we be doing? How should we refine our strategy going forward?

And, while most program evaluations are geared toward use by program staff, strategic evaluations should also serve senior-level leaders and boards, as they consider the future of the program or initiative and course corrections, refinements, and/or enhancements to the organization's strategy(ies).

I also continue to believe that the most informative and useful evaluations are those that are adaptive, flexible, nimble, responsive, occur in real time, collaborative, and systems oriented (Preskill & Gopal, 2014). As organizations and communities strive to find new ways to solve social problems, evaluators must also evolve and grow in their approaches to increase not only the quality of the data and findings we deliver but also the ways in which we do our work. This includes expanding the evaluator's toolkit to include techniques and approaches to facilitating real-time learning around data and experiences to engage stakeholders in sense-making opportunities around findings and recommendations. To this end, my colleagues and I developed the *Facilitating Intentional Group Learning* guide (Preskill, Gutiérrez, & Mack, 2017), as a means to further embed learning practices into our daily work.

Over the last decade, I have also deepened my understanding of what it means to engage in positive systems change, and how evaluation can contribute to this. Defining *systems change* as “changing the conditions that are holding the problem in place” (Kania, Kramer, & Senge, 2018; Social Innovation Generation), evaluators need to consider what those conditions are—for example, policies, practices, resource flows, relationships and connections, power dynamics, and mental models—and how best to evaluate them. But a topic that has been rarely discussed in the systems change and complexity literature is how racism and structural inequities across populations of color, ability, sexual orientation, and gender have been root causes of maintaining the status quo of systems. When we talk about evaluating the conditions that are holding problems in place, no longer can we ignore racism as a root cause. I've been fortunate to work in an organization that has been on a multiyear journey in learning about racial equity and racial justice, and, more recently, has been engaged in our own reckoning concerning the ways in which we unintentionally may have contributed to continu-

ing social inequities, internally and through its client engagements and writings. The disparities faced by Black and Brown communities—as surfaced by the impacts of COVID-19 and documented police killings of Black men and women—have become highly visible and undeniable, precipitating a social uprising for racial justice unprecedented in the speed at which it has gained broad, national attention. Evaluators have been called on to question evaluation orthodoxies and how we (particularly White evaluators) have been complicit in perpetuating these injustices and inequities through our work (e.g., Jara Dean-Coffey, *Equitable Evaluation*; www.equitableeval.org).

For more than 20 years, brilliant evaluation scholars and practitioners (e.g., Rodney Hopson, Stafford Hood, Veronica Thomas, Mel Hall, Joan LaFrance, Donna Mertens, and others) have been writing about issues related to cultural awareness, cultural competence, and equitable evaluation practices. But today seems different, in that a confluence of historic events has accelerated our questioning about what good evaluation looks like; for whom we evaluate; who has power with, in, and over; and who benefits from evaluation processes and findings (as well as who might be harmed). My hope is that we can commit to finding ways of centering equity in our evaluation work, in ways we've never done before.

These three phases of my career have contributed to how I've thought about and practiced evaluation over the last four decades. Though it's difficult to sum it all up, I do think the following assumptions reflect the specifics of my evolving evaluation theory.

Assumptions about the Purpose of Evaluation

- Evaluation is a catalyst for individual, group, organization, and community learning. As such, it is change-oriented and constitutes an intervention at whatever levels it is conducted.
- Evaluations should be conducted only when there is an intention to use

the findings and includes instrumental, conceptual–enlightenment–knowledge uses and political–symbolic–persuasive uses.

- Evaluations should focus on questions that matter; they should reflect the strategic information needs of the organization or community, and be designed to inform understandings, decision making, and potential actions.
- Expected uses of evaluation processes and findings should guide an evaluation's design and implementation.
- Evaluation should contribute to the social good. They should include questions about power, privilege, structural inequities, values, and resources.

Assumptions about the Nature of Evaluation Practice

My assumptions about evaluation practice reflect a belief that evaluation is most effective, meaningful, and useful when it is conducted using collaborative, participatory, learning-focused, and systems-oriented approaches that pay particular attention to racial equity and justice.

- Evaluation should be ongoing, reflexive, and embedded in organizational practice. As such, evaluative inquiry is an open-ended, continuous process that is integrated into the everyday work of organization members. It does not constitute an event, nor is it product-driven. Instead, it seeks to address critical organizational issues and questions as they arise in the course of one's work. This means that all organization members are responsible for thinking evaluatively—for understanding the role of evaluation and how and when it might be used to gather information to inform decision making and action.
- When a diverse set of stakeholders is involved in various phases of an evaluation, stakeholders themselves not only learn more about evaluation but the evaluand and the intended users of the evaluation findings are also more likely to use them.

- Involving a wide range of stakeholders increases the likelihood that multiple perspectives and life experiences and values will guide the evaluation's design, as well as the interpretation of the data, thus resulting in more diverse and inclusive findings.

- Involving stakeholders in evaluation processes contributes to building their own capacity to engage in evaluative thinking and practice.

- Learning from the evaluation process is an important goal (i.e., process use). This means providing space and time for asking questions; engaging in dialogue; identifying and challenging values, beliefs, and assumptions; and reflection (both public and private). Being clear about the objectives of learning and using a variety of learning activities enhances evaluative capacity building.

- The most effective evaluator role is that of facilitator, guide, educator, mentor, critical friend, instigator, and provocateur.

- Evaluators who use learning approaches that are collaborative and participatory are more effective if they understand the concepts of group dynamics, systems change and complexity theory, trust and power, organizational change and culture, adult learning, self-efficacy, racial justice, equitable evaluation, and culturally aware practices. They are also more successful if they are able to facilitate meetings effectively, provide feedback, listen actively, mediate conflict, and negotiate compromise.

Assumptions about the Context of Evaluation Work

My assumptions here reflect the belief that evaluation occurs within dynamic, unpredictable, and nonlinear environments. As such, evaluators need to recognize the interconnections of various parts of the systems in which they work, the inherent structural inequities that are built into systems, and the political and value-laden nature of their

work. In addition, when working in organizations, evaluation practice and the use of evaluation findings are always influenced by the organization's culture, systems and structures, and leadership behaviors.

- Evaluation in service of systems change focuses on questions concerning the conditions that are holding problems in place. Learning-oriented evaluators seek to understand the ways in which policies, practices, resource flows, relationships and connections, power dynamics, mental models, and racism influence systems change. Evaluators' work includes the exploration of how and why different components and sub-systems interact (or do not) in producing the desired outcomes. Understanding and describing the system in its entirety, including specific components and connections, is an integral part of an evaluation.

- Evaluation is inherently a political activity—that is, it is about power. Consequently, evaluators must consider the political influences and environment when negotiating, designing, and implementing an evaluation. Among other things, these considerations might affect the involvement of certain stakeholders, possible ethical conflicts, and/or the potential use or misuse of evaluation processes and findings.

- How an organization is structured, how individuals' jobs are designed, how supportive the organization's culture is of evaluative inquiry, the extent to which leadership models and supports learning and inquiry, as well as how information is distributed and accessed, strongly influence the extent and the ways in which evaluation is successful.

Assumptions about Developing Evaluation Recommendations and Communicating Evaluation Findings

The following assumptions reflect a long-held belief that evaluators are missing opportunities to enhance learning toward the end of an evaluation study. The ways

in which we engage in *sense making*—developing interpretations, judgments, and recommendations—and how we communicate these findings (i.e., learnings), can positively influence evaluation use.

- Engaging stakeholders in collective sense-making activities with analyzed data and using intentional learning activities can accomplish two important goals: (1) deepen the diversity, accuracy, and relevance of the resulting considerations or recommendations and (2) build the capacity of stakeholders' evaluative thinking.

- Different formats for communicating evaluation processes and findings are essential for enhancing audiences' understanding of the information and the extent to which it is useful and used.

- Various audiences have different information needs and therefore require different forms and frequencies of communication.

INFLUENCES ON MY POINT OF VIEW

I have learned so much from so many over the years. In fact, I think that one of the reasons I am still passionate about evaluation is that it has always stimulated me to keep learning. Conducting a wide variety of evaluations and research on evaluation, teaching evaluation, mentoring evaluation students, and working with philanthropic, corporate, and nonprofit organizations seeking to make evaluation meaningful always raises new questions and issues that continue to shape my evaluation point of view. At the same time, there is little doubt that several scholars in various fields have significantly influenced my theory of evaluation.

From Bob Stake, I learned about the importance of responding to stakeholders' information needs, the value of representing the multiple realities of program participants, and the art of the case study.

The early work of Marv Alkin, Michael Q. Patton, Jean King, and Carol Weiss on evaluation use not only helped me focus my dissertation research but also paved the way for making evaluation use the focus of my long-term research agenda and consulting practice. And Patton's critically important writings on process use, facilitation, and developmental evaluation have motivated me to explore new ways of thinking about the evaluator's role and how evaluation can foster individual, team, and organizational learning.

Many individuals have influenced my belief in the value of participatory and collaborative approaches to evaluation. Jennifer Greene's candid and reflective writings on the realities of participatory evaluation have forced me to be more truthful and honest about the realities of involving stakeholders in evaluation work. Brad Cousins's work (with Lorna Earl, Lyn Shulha, and Elizabeth Whitmore) has helped me understand the nuances of participatory evaluation and the challenges individuals face in doing such work.

Though my approach to evaluation didn't always address the politics of evaluation explicitly, I am acutely aware of the political nature of evaluation and the issue of power (e.g., who has it, how it is wielded, who it excludes) that is inherent in all evaluation approaches. The political uses of evaluation, social justice, and democratic deliberation, as Carol Weiss, Ernie House, and Sandra Mathison have respectively explored, always seem to be in the back of my mind. Their work has kept me from being too idealistic and naive while still remaining hopeful and optimistic. As I reflect on where the world is today and, in particular, the United States, I believe evaluators need to understand different kinds of power and how they contribute to maintaining the status quo (e.g., structural inequities). We need to continue asking questions about power—who has it, how it is being wielded and to what ends and with what consequences, and how it can be redistributed and shared.

As my work increasingly took on an organizational learning focus, I read the work of Chris Argyris, Donald Schon, Peter Senge, Karen Watkins, and Victoria Marsick. Their writings affirmed the idea that evaluation, being an approach to organizational inquiry, could help organizations resolve critical issues and grow from the collective experiences of their members. I found that the approaches to team and organizational learning, which were based on social constructivism, were also related to participatory and collaborative forms of evaluation. And, from the Organizational Learning literature, I learned about Appreciative Inquiry (AI). As I explored this philosophy and approach to organization, community, and social change, I was inspired to find ways to bring it into the evaluation field, and my work specifically. I am greatly indebted to Tessie Catsambas, who was willing to share her rich insights and experiences that culminated in our book, *Reframing Evaluation through Appreciative Inquiry* (Preskill & Catsambas, 2006). AI is a process that inquires into, identifies, and further develops the best of “what is” in organizations. It is best described as an approach, framework, or mind-set that focuses on illuminating and affirming personal success factors or forces within an organization, to be used with existing organizational development interventions such as strategic planning, organizational design or restructuring, and project evaluations (Watkins & Mohr, 2001). Shifting from a deficit-based model and language that reflects most evaluation work to a more appreciative, positive-based model language is extremely appealing. The work of David Cooperrider, who developed this approach in the 1980s along with others, including Diana Whitney, Jane Watkins, and Bernard Mohr, helped us connect AI’s philosophy and practice with participatory, collaborative, and learning-oriented evaluation approaches.

The third phase of my career has been influenced by an array of people I greatly admire. Tanya Beer and Julia Coffman are

two of them; their work on strategic learning, the role of learning in philanthropy, and learning biases, among other topics, have continually informed and inspired me. Their generous spirit and willingness to push the field’s thinking and practice have often kept me going. In the field of complexity and systems thinking and change, I’ve learned so much from Brenda Zimmerman, Bob Williams, Glenda Eyong, Srik Gopal, Beverly Parsons, and Meg Hargraves. There are many who write on this topic, but these folks in particular have helped make understanding systems change more understandable and actionable. Their writings, workshops, and collegial conversations have often been the accelerant to my evolving understanding of how systems change happens (or does not). Finally, I have learned immeasurable amounts from my FSG colleagues, who continue to wonder, question, collaborate, and share their learnings with me; I will be forever thankful for their generosity.

Finally, these last few years have also been a period of deep reflection and growth when it comes to understanding racial equity, structural inequities, and social justice. The workshops and writings of Jara Dean-Coffey, Robin D’Angelo, John A. Powell, and organizations such as Proinspire, Frontline Solutions, and Change Elemental, among others, have influenced how not only my evaluation theory has evolved, but how I, as a White woman, am showing up as an evaluator, as a citizen, and as a human being. I am grateful for their perseverance, honesty, vulnerability, and courage, in challenging us to be better.

PARTING THOUGHTS

As always, writing a chapter such as this provides a special and rare opportunity to reflect on where one has been and where one wants to go. It has allowed me to reflect deeply on my values, experiences, hopes, and desires for evaluation. Writing about

my evaluation journey has enabled me to reaffirm my convictions about the value of evaluation, my passion for evaluation, and the possibilities of evaluation, topics near and dear to my heart and mind. Yet I also realize that putting such thoughts on paper has its downside. It seems to fix things in time, when, in reality, theories of any kind are always in progress, always being tested and revised, as Alkin and Christie remind us. Though aspects of my evaluation point of view will likely continue to evolve in the coming years, I do know that I will always be committed to making evaluation processes and outcomes useful, and will do everything I can to enhance what may be learned from evaluation practice.

REFERENCES

- Kania, J., Kramer, M., & Senge, P. (2018). *The water of systems change*. www.fsg.org/publications/water_of_systems_change
- Preskill, H., & Boyle, S. (2008). A conceptual model of evaluation capacity building: A multidisciplinary perspective. *American Journal of Evaluation*, 29(4), 443–459.
- Preskill, H., & Catsambas, T. (2006). *Reframing evaluation through appreciative inquiry*. SAGE.
- Preskill, H., & Gopal, S. (2014). *Evaluating complexity: Propositions for improving practice*. www.fsg.org/publications/evaluating-complexity
- Preskill, H., Gutiérrez, E., & Mack, K. (2017). *Facilitating intentional group learning*. www.fsg.org/tools-and-resources/facilitating-intentional-group-learning
- Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. SAGE.
- Preskill, H., & Torres, R. T. (2000a). The learning dimension of evaluation use. *New Directions for Evaluation*, 88, 25–38.
- Preskill, H., & Torres, R. T. (2000b). *Readiness for Organizational Learning and Evaluation Instrument (ROLE)*. Available from the authors.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2), 5–8.
- Torres, R. T., Preskill, H., & Piontek, M. E. (1996). *Evaluation strategies for communicating and reporting*. SAGE.
- Watkins, J., & Mohr, B. (2001). *Appreciative inquiry: Change at the speed of imagination*. Jossey-Bass/Pfeiffer.

CHAPTER 20

The “Theory” of Interactive Evaluation Practice

Jean A. King

INTRODUCTION

As students in my introductory evaluation course learn, I firmly believe that everyone is an evaluator. We are all evaluators because we make judgments on a daily basis, whether determining the best route home in an ice storm, deciding which new car to lease or perhaps purchase, or—and here’s where we come in—capturing the process and outcomes of a program to judge its effectiveness or to make decisions about how best to move forward. I see our job as evaluators to help people systematize something that is a routine part of their lives, even if they’ve never thought about it before. When I am asked how to structure an evaluation, the specific details of a program and the contextual diversity often lead me to the classic evaluator response of “It depends” because the best answer (or even one that is good enough) really does depend on multiple things—the setting, people’s value commitments, resources available, a program’s history, and so on. Thus, I have learned over

the years that, while people routinely engage in a kind of evaluative thinking when they make decisions, the process of program evaluation that Scriven (1980) explains so clearly in his logical frame is nevertheless complex.

This brings me to the concept of evaluation theory. As I have written before (King, 2012), I have never considered myself an evaluation theorist. Rather, I am someone who practices evaluation, an endeavor in which practice can never make perfect for at least two reasons: the competing values of the clients who engage us and the multiplicity of contexts in which we work, even within a single organization. When I was in graduate school in education in the 1970s, I took several courses devoted to the development of explanatory theories. These included studying the methodologies that would lead a researcher to ultimately illuminate some partial truth about whatever it was he or she was studying, be that, for example, the effect of English teachers’ comments on high school students’ papers (my disserta-

tion topic), the efficacy of teaching young children conflict management techniques, or the role of parental involvement in students’ academic achievement.

When I became an evaluator, I was surprised to learn that the process for creating theory I had learned in my doctoral study was not necessarily the one that my newfound field applied. Instead, our field contrasts such traditional forms of social science theory, labeled descriptive theory, with the so-called “prescriptive theory” of evaluation practitioners (Alkin & House, 1992). By this definition, then, I learned that I, too, am technically an evaluation theorist because over time I have developed a consistent approach to my practice that I find helpful. Eventually called *interactive evaluation practice* (IEP), it was explicated in collaboration with my longtime coauthor and friend, social psychologist Laurie Stevahn (King & Stevahn, 2013; Stevahn & King, 2015).

THE WHAT, WHERE, AND WHO OF IEP

This section begins with an explanation of what IEP is, then describes where it is most likely to be successful and the attributes of successful IEP practitioners.

What Is Interactive Evaluation Practice?

King and Stevahn (2013) define IEP as “the intentional act of engaging people in making decisions, taking action, and reflecting while conducting an evaluation study” (p. 14). It is grounded in the *interpersonal factor*, which “highlights the unique ability of an evaluator to do two things: (a) interact with people constructively throughout the framing and implementation of evaluation studies and (b) create activities and conditions conducive to positive interactions among evaluation participants” (p. 6). It is important to notice that IEP is not only about *beginning* a study (i.e., “breaking the ice”) and initially getting people to interact

with one another positively. At its best, it is part and parcel of the entire evaluation process, start to finish—and even beyond.

In my opinion, IEP builds on utilization-focused evaluation (U-FE; Patton, 2010, 2012) in that it can be fitting for any type of evaluation, even mandated accountability studies or those that require strict forms of quantitative methods. IEP’s purpose is to foster the constructive engagement of “primary intended users” (i.e., critical stakeholders who will eventually act on the results), both to make the study technically sound and to build people’s interest and potential use into the process. The approach is highly participatory and time-consuming. For this reason, I find that an ideal setting for this approach—although not required—includes a long-term engagement with an organization willing to commit to the extensive involvement of groups of collaborators over time.

The Goals of IEP

IEP is a participatory process that seeks to engage key stakeholders meaningfully in the entire evaluation process. It has three ultimate goals, which unavoidably overlap:

1. Teach decision makers to engage collaboratively so that, over time and to the extent possible, the evaluation process and use of its outcomes become integrated into the program’s or organization’s ongoing functioning. In so doing, IEP works to ensure that the evaluation process addresses the right issues (i.e., those important to that setting and that matter to those involved in the study), the methods used are appropriate and grounded in stakeholders’ extensive knowledge of the context and its participants, and data analysis and interpretation are demystified so that key participants both own the findings and understand them well.

2. Ideally—and this obviously doesn’t always happen—a second IEP goal is to help leaders enhance their personal evaluation

skills to the extent possible and, as the first goal indicates, to develop their organization's capacity to do and to use evaluation. I have written elsewhere (King, 2012) that the highest form of evaluation for me is called *free-range evaluation*, a process that lives independently in an organization and has to be tenacious to survive. Merely by participating in an IEP study, people may learn evaluation skills, systematize the intuitive process they know, and come to value the process leading to free-range evaluation. In IEP, American Evaluation Association (AEA) competency 4.9 ("Works with stakeholders to build evaluation capacity when appropriate") is always a consideration.

3. Because the process is highly interactive, a final goal is to support a more positive environment and functional working relationships as a result of the conversations and interactions people have. Those who participate may actually enjoy interactions around evaluation and are often surprised when that happens.

Settings Appropriate for IEP

I would never insist or recommend—which the term *prescriptive* theory (Alkin & House, 1992) suggests—that anyone use IEP in settings where it does not fit. It may, by now, be obvious to say that context is key (Rog, Fitzpatrick, & Conner, 2012). At its best, program evaluation is context-specific, and the "theory" that I describe here fits better in certain contexts than in others. In considering IEP, the power of that assertion cannot be overstated. But it should also be noted that whereas IEP may only work well in certain sites, it is nonetheless a general approach that applies not only to programs with specific content (e.g., smoking cessation or increasing crop production) but also to any program. The task finally is to identify settings where IEP makes sense and interact evaluatively with the individuals who are living and acting in them.

What do IEP-appropriate settings look like? Because the intent is to engage people meaningfully throughout the evaluation process, in my experience, it works well in relatively small programs or in programs that have multiple sites in which it is possible to interact with virtually every essential stakeholder (i.e., all of the key people who are willing and able to help frame the study), decide how to collect data, assist with data collection and interpretation, and finally, figure out how to use the evaluation's results. As mentioned earlier, a long-term engagement can help an evaluator structure multiple interactions that over time lead to evaluation capacity building.

Clearly, there must be an openness to the evaluation process and a readiness to accept its outcomes. IEP cannot work in settings where distrust is pervasive or where people in leadership roles refuse to relinquish or share their power. If recommendations that emerge cannot be acted on, either because people lack the authority themselves or because a superior refuses to, the process is likely to frustrate everyone who took part. This description of appropriate IEP settings suggests why it is *not* appropriate for many studies. I must further note that the COVID-19 pandemic has definitely added an exceptional, although likely necessarily insurmountable challenge to group processes since they may no longer occur in person. Zoom is a helpful and cost-effective alternative, but to me, it will never be a substitute for face-to-face interactions.

Prerequisites for IEP Practitioners

While the setting matters to IEP, so, too, do the evaluators choosing to implement the approach. A critical prerequisite for IEP, literally a *sine qua non* for its successful use, is a finely honed set of interpersonal skills. Let me say it again. Absent high-quality interpersonal skills, IEP will surely fail. Donaldson, Gooler, and Scriven (2002, p. 261) write that "more than technical skills are required to

conduct high-quality program evaluations," but to date, the field of evaluator education has not focused extensively on the formal, systematic development of interpersonal skills (King & Ayoo, 2020). In the 2018 AEA Evaluator Competencies, the description for the interpersonal domain states that the domain's seven competencies focus "on human relations and social interactions that ground evaluator effectiveness for professional practice throughout the evaluation. Interpersonal skills include cultural competence, communication, facilitation, and conflict resolution" (AEA, 2018). Each of these skills matters in IEP, and evaluators interested in using the approach should make sure that they or their team members collectively embody all of them.

Features of the IEP skills set include an ability to put people at ease, to listen thoughtfully (especially to those directly affected by a program), to read and confirm subtle nonverbal cues, to ask open-ended questions, and to identify and manage conflict productively. Participants who are either ignoring or actively undermining the evaluation need to be engaged in ways that, if possible, will get the process back on track. Facilitation skills are part of the interpersonal requirements, including an ability to facilitate "Constructive and culturally responsive interaction throughout the evaluation" (AEA competency 5.7). Can people learn or refine these skills if they lack confidence in them? I believe they can. As comparison, consider the training of clinical psychologists, social workers, or therapists in which students routinely work to develop these skills. The specific techniques and actions outlined below may provide an outline for their development.

To my mind, participation in an evaluation should be an educational experience, with the evaluator serving as a thoughtful and attentive teacher (Cronbach & Associates, 1980). You need figurative eyes in the back of your head and at least some participants who are willing to share what is really

going on. In my practice of IEP, there are three other prerequisites: (1) sincere respect for the context and people with whom you interact; (2) humility, which relates to awareness of yourself both as a person and as others perceive you, especially in the sometimes frightening or off-putting role of evaluator; and (3) an unfailing sense of humor. One thing is certain: Whereas everyone else can whine, complain, and get annoyed, an IEP evaluator must maintain an upbeat attitude throughout. Writing about evaluation in indigenous communities, Bowman (2020) suggests that evaluators remain "unapologetically positive," which fits well with IEP. In keeping with King Rule for Living #4 (Stevahn & King, 2010, p. 154), then, it is important for IEP evaluators to always be nice. This is sometimes easier said than done.

THE PRINCIPLES OF IEP

Part I of King and Stevahn (2013) provides research grounding for seven principles underlying interactive evaluation practice.¹ In the interest of word count, I have integrated these principles into two overarching ideas, which I briefly explicate: (1) get personal and (2) structure interaction purposefully.

Get Personal

IEP is built upon relationships, so getting personal (i.e., getting to know people—well) and interacting with them productively is a grounding principle. Every evaluation setting is unique in time and place, and it is critical to engage a diverse group of positional leaders, their informal counterparts, key opinion leaders, and even naysayers over time, fostering positive relationships that increase trust in both you and in the evaluation process. I know that the people with whom I interact have their own knowledge and experiences that they actively bring to bear during the evaluation process, and it

is important that they do so. Therefore, IEP cannot be rushed. For example, at the most basic level, the following takes time:

- Schedule meetings, which are sometimes delayed because evaluation may not be a top priority for people you need to include.
- Conduct sessions with *all* key participants present in both body and mind; higher-ups often want to give introductory hortatory comments and leave, presumably for other more important activities.
- Prepare data for interactive analysis once available.
- Ultimately, frame findings and plans for next steps.

An ability to understand the organizational context, including its dynamics, politics, and cultural concerns, is paramount. As the 2018 AEA Evaluator Competencies detail, all evaluators are responsible for “Clarify[ing] diverse perspectives, stakeholder interests, and cultural assumptions” (AEA competency 3.7), “Understand[ing] and engag[ing] different perspectives” (AEA competency 5.2), and “Facilitat[ing] shared decision making” (AEA competency 5.3). IEP elevates the importance of these competencies, and to do these well requires connecting on a personal level. By establishing a high degree of trust, an evaluator can build relationships able to weather or at least address the challenges that will inevitably occur. Doing so when people are unable to gather in person (e.g., using Zoom or FaceTime) may prove difficult, if not impossible.

Structure Interaction Purposefully

A second principle requires IEP evaluators to actively plan and implement interactions, large and small, between and among evaluation participants, a definite challenge in the context of the COVID-19 pandemic. This principle highlights AEA competency 5.7: “Facilitates constructive and cultur-

ally responsive interaction throughout the evaluation.” Always done in collaboration with key clients, this includes scheduling gatherings, building agendas, designing interactive activities, facilitating discussions and meetings, and whatever else the process needs to move forward. Some derisively label such extensive interaction as overly “touchy feely,” but in my experience, it is these activities that build the relationships that sustain the evaluation process and foster the eventual use of its findings. One aim is to enable “shared understanding of the program and its evaluation with stakeholders” (AEA competency 3.6). IEP requires evaluators to accept responsibility for encouraging use by enacting key U-FE activities: framing and prioritizing questions, agreeing on methodology and reviewing instrumentation, and ultimately engaging in data analysis and developing recommendations. In this way, IEP enacts AEA competency 3.8 (“Promotes evaluation use and influence in context”).

Knowing the likelihood of conflict in evaluation, IEP practice also requires evaluators to expect it, watching carefully when people interact, and focusing on how those with power in the setting engage with the evaluation process. Structuring interaction includes attending to and managing any conflicts that arise (AEA competency 5.7). High-quality and transparent communication is essential, remembering that the evaluator’s power is indirect, typically stemming from ongoing relationships with those who commissioned the study (King, 2016). When conflicts do arise, they require direct attention, which can be truly challenging—or even impossible—in organizations where this is not the norm. Knowing when a conflict has derailed the evaluation process, an IEP evaluator can either seek integrative negotiation or a compromise (see King & Stevahn, 2013, Ch. 6)—or, as a last resort, choose to end the study. Because the ways that people interact matter, an evaluator’s ability to structure effective interactions and manage or resolve conflict are important skills that can lead to successful IEP.

BRINGING INTERACTIVE EVALUATION PRACTICE TO LIFE

How, then, do you implement interactive evaluation practice? And how would you know an IEP study if you saw it? Paralleling the roles for participants included in IEP's definition, King and Stevahn (2013) identify the same three roles for an evaluator: decision maker, actor, and reflective practitioner. Applying these roles, IEP requires data-based decisions, planful actions, and continuing reflection. In important ways, IEP is a form of developmental evaluation (Patton, 2011) because the process evolves in response to the evaluator's reflection on events as they occur, modeling AEA competency 1.5 ("Reflects on evaluation formally or informally to improve practice").

The assignment for this chapter was to provide information to "translate (interactive evaluation practice) . . . into a working guide for implementation." That is, finally, an impossible task in the space available. The five chapters in Part II of King and Stevahn (2013) detail specific "Skills and Strategies for Interactive Evaluation Practice," dividing the content into how to conduct evaluation conversations, how to facilitate productive interactions, how to manage conflict constructively, how to create a viable interactive process overall, and how to deal with the unexpected. Given the brief length and descriptive nature of this chapter, I outline some of the activities and refer those who are interested to the book for a more complete how-to.

First, IEP practitioners need to know how to have productive conversations with a variety of people because, absent frequent conversations, evaluators cannot do their job. Three types of concerns are included in these conversations: technical, political, and resource-related. IEP highlights the need to have open discussions with key participants both before and after agreeing to conduct a study and then during the study and afterward. Take, for example, a conversation about how a proposed study is to

be conducted. There are several topics to discuss, among them the overall approach (purposes and potential uses), the overarching questions the study seeks to answer, the information required, a feasible design, and credible methods, among others. There are also potential pitfalls, including clients who are unclear about what they actually want, who don't want to participate in the conversation, or who lack sufficient resources for the study (King & Stevahn, 2013, p. 79). This initial discussion would undoubtedly require numerous follow-ons as the study progresses. A successful IEP evaluator must learn to converse effectively with participants throughout the evaluation, including, as noted previously, a process for resolving or managing conflicts.

Second, as stated earlier, IEP requires evaluators to structure interactions that enable participants to connect with one another for evaluative purposes. There are many facilitation techniques for such engagement (e.g., Chaplowe & Cousins, 2016; Fierro, Schwartz, & Smart, 2016; Preskill & Russ-Eft, 2016), some of which can be adapted for virtual online use. Among the 13 strategies included in Chapter 5 (I'm choosing those that I find especially helpful) are what we call voicing variables, choosing corners, cooperative interviews, cooperative rank order, and dot votes/bar graphs. Because IEP is grounded in social interdependence theory, these strategies build on cooperative goal structures, promotive interactions, and mutually beneficial outcomes. The cooperative (three-step) interview, for example, has groups of three people working together. They share a question sheet; take turns alternately answering the question, asking follow-ups, or recording the answer; and then together analyze the data gathered, looking for common ideas and differences in their answers. The technique can work with groups large or small; everyone present is divided into groups of three, and each group often has a powerful conversation.

Finally, IEP identifies five components for planning and managing a viable interactive

evaluation: (1) knowing the context well; (2) understanding two types of “culture” in that context (i.e., organizational culture and that related to issues that the cultural competence/social justice literature has addressed); (3) determining who will be involved in the process; (4) organizing logistics; and (5) communicating the process and its results. Ongoing evaluative conversations and carefully structured interpersonal interactions move each of these steps forward. Over time, I have learned the critical importance of understanding culture in relation to the “dynamics of power” that the AEA Statement on Cultural Competence in Evaluation (2011) names, which is the topic of the following section.

THE CHALLENGE OF POSITIONALITY IN IEP

In the decade or so since the second edition of *Evaluation Roots* was published, thankfully, the field has moved toward a greater awareness of our role in inadvertently supporting a powerful social infrastructure that sustains racial and economic injustice in this country. The events of 2020—the pandemic, George Floyd’s murder, and the U.S. presidential campaign—sadly highlighted the dramatically dissimilar experiences different people have in this country, depending on who they are, what they look like, and where they live. The “dynamics of power” obviously affect many people negatively. As an IEP evaluator, I must assess my position as it relates to this social and political context, and for me, that is a challenge. What exactly do I, as a person, bring to my evaluation practice at this difficult moment in history?

For well over 20 years, I have been keenly aware of my privileged position as a person, a professor, and an evaluator (King, 1998). As a professional evaluator seeking to repair the world (*tikkun olam* in Hebrew), I identified a personal dilemma: the Mother Teresa versus Princess Diana problem. Both women sought to help people in need, Mother Teresa

by working directly with those who struggled desperately in poverty, and Princess Diana by committing herself to good causes like land mine reform and reducing the stigma of HIV/AIDS. But Mother Teresa lived a life of squalid hardship, while—even though we now better understand the stresses with which she struggled—Princess Di lived a life of incredible privilege. As a White, upper-middle-class evaluator who earns a good salary as a professor and consultant, I live a lifestyle that is closer in kind to Princess Di’s than to Mother Teresa’s, which makes me feel guilty. I have never worried about the details of basic survival, my family life has been stable for over 70 years, and until the pandemic, I frequently traveled far and wide. I am not anxious about my children being stopped by the police when they are driving. Why do *I* have this privilege? And knowing my privilege, why do I think I can work effectively in any community that is willing to engage with me?

The fact that I am a Jew, a member of a minority that anti-Semites are increasingly targeting in this country and abroad (although Jews are well less than 1% of the U.S. population), adds somewhat to the complexity of my position. Anti-Semites may blame Jews for controlling Hollywood and the rest of the world, but as far as I know, my being Jewish has never limited my options either personally or professionally. I don’t “look Jewish,” and I don’t have an overtly Jewish surname. Furthermore, the minority of which I am part has been able to succeed in this country, in part, yes, not only because the culture values education and hard work, but also because, unlike members of many other minority groups, we can ultimately pass in White society. This is another inequity that makes me feel guilty, and to me, this means that I must be even *more* reflective about the ways that my positionality may affect the evaluation process in a given context.

Because IEP requires you to interact effectively with community members, you need to think hard about whether, given who you

are, this is even possible. Good intentions may be necessary, but they are surely not sufficient for privileged evaluators. One of my most important learnings about potential relationships came from an Aboriginal Australian elder, a woman who had participated actively in a participatory evaluation project. She had been invited to speak about her experience at an Australasian Evaluation Society conference, and a bright-eyed young evaluator (not me) asked her about how well the process could work. My memory of her response goes like this:

"I think about my community's knowledge as a cake. It has three layers. The outside layer is what everyone can see—the obvious facts and details, things that people can look at or read about. Inside that is the second layer, which you may be able to learn about if you come to our community, speak respectfully with us, and spend enough time to truly get to know us. But the very inner part of the cake, the third and most important part, belongs to us. It is not yours to know about, and we will never reveal it to you."

What this means for me is that we evaluators can definitely access the cake's first layer, and, with care and time, we may come to understand the second. But it is important to accept the fact that, except in our own communities, we do not have a right to access the third. As the elder said, it is not ours to know. It is the community members' privilege to keep it to themselves.

This leads me finally to three thoughts for how IEP evaluators can respond to concerns about positionality.

1. Centuries ago, Socrates wrote that "to know thyself is the beginning of wisdom." The first action, then, is to commit to learning about yourself and truly understanding how you fit into the big picture of the world in which you live. How do you relate to the "dynamics of power" in the cultural, historical, and political systems (large and small)

of which you are a part? What privilege do you have? How does that affect your evaluation practice? What power *do* you have to take meaningful action?

2. It is important to ask yourself whether, in light of your self-knowledge, you are the right person to engage a community or a group of people in an interactive evaluation process. You must sometimes acknowledge that there are settings in which you simply cannot lead an interactive study, although you perhaps can coach and support someone who is more appropriate for the task.

3. If you are interested and committed, IEP may lend itself to beginning the process of decolonizing evaluation. Several of IEP's grounding principles appear to parallel the "principles of liberating structures" (e.g., "practice deep respect for people and local solutions," "build trust as you go," "engage in seriously playful curiosity"), and the right evaluator may be able to employ IEP strategies that support community members in conducting useful evaluations and building evaluation capacity.

CONCLUSION

This chapter has summarized the work of a lifetime, an approach to evaluation that I call interactive evaluation practice. The goals and two overarching principles clarify IEP's intentions, and previous sections have identified the kinds of settings in which I've found success and the prerequisites I believe IEP evaluators require. The brief discussion of how to implement IEP points to its distinguishing features. An IEP evaluation engages key individuals in ongoing conversations about every aspect of the study and, from beginning to end, creates numerous options for people's interaction, whether in person or virtually. As the field has evolved over my career, so, too, has the approach, with an awareness now that it may prove useful in building people's capacity to use evaluation to strengthen their communities.

I sincerely hope that this can and will happen in the coming years.

ACKNOWLEDGMENT

I want to thank Laurie Stevahn for her thoughtful suggestions on improving this chapter.

NOTES

1. The seven IEP principles are get personal, structure interaction, examine context, consider politics, expect conflict, respect culture, and take time (King & Stevahn, 2013, p. 61).

REFERENCES

- Alkin, M. C., & House, E. (1992). Evaluation of programs. In M. C. Alkin (Ed.), *Encyclopedia of educational research*. Macmillan.
- American Evaluation Association. (2011). *American Evaluation Association Statement on Cultural Competence in Evaluation*. www.eval.org/p/cm/ld/fid=92
- American Evaluation Association. (2018). *AEA competencies*. www.eval.org/p/cm/ld/fid=472
- Bowman, N. (2020). Asking questions assuming normal. Listserv, April 11, 2020, EvalTalk. <https://listserv.ua.edu/archives/evaltalk.html>
- Chaplowe, S. G., & Cousins, J. B. (2016). *Monitoring and evaluation training: A systematic approach*. SAGE.
- Cronbach, L. J., & Associates (1980). *Toward reform of program evaluation*. Jossey-Bass.
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation*, 23(3), 261–273.
- Fierro, R. S., Schwartz, A., & Smart, D. H. (Eds.). (2016). Evaluation and facilitation. *New Directions for Evaluation*, 149, 31–42.
- King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation*, 80, 57–67.
- King, J. A. (2012). Getting people involved: The origin of interactive evaluation. In M. C. Alkin (Ed.), *Evaluation roots* (2nd ed.). SAGE.
- King, J. A. (2016). Taking what action for change? Dare evaluation build a new social order? *Visitors Studies*, 19(1), 3–11.
- King, J. A., & Ayoo, S. (2020). What do we know about evaluator education?: A review of peer-reviewed publications (1978–2018). *Evaluation and Program Planning*, 79, Article 101875.
- King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. SAGE.
- Patton, M. Q. (2010). *Utilization-focused evaluation*. SAGE.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. SAGE.
- Preskill, H., & Russ-Eft, D. (2016). *Building evaluation capacity: Activities for teaching and training* (2nd ed.). SAGE.
- Rog, D. J., Fitzpatrick, J. L., & Conner, R. F. (Eds.). (2012). Context: A framework for its influence on evaluation practice. *New Directions for Evaluation*, 135, 1–105.
- Scriven, M. (1980). *The logic of evaluation*. Edgepress.
- Stevahn, L., & King, J. A. (2010). *Needs assessment phase III: Taking action for change*. SAGE.
- Stevahn, L., & King, J. A. (2015). Facilitating interactive evaluation practice: Engaging stakeholders constructively. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and facilitation. New Directions for Evaluation*, 149, 61–74.

CHAPTER 21

Developmental Evaluation

Michael Quinn Patton

INTRODUCTION

Developmental evaluation (DE) provides *evaluative* information and feedback to social innovators and their funders and supporters to inform adaptive *development* of change initiatives in complex dynamic environments. DE brings to innovation and adaptation the processes of asking evaluative questions, applying evaluation logic, and gathering and reporting evaluative data to inform and support the development of innovative projects, programs, initiatives, products, organizations, and/or systems change efforts with timely feedback.

THE NICHE OF DE

The DE niche focuses on evaluating innovations in complex dynamic environments because this is the area in which *social innovators* are working. Innovation as used here is a broad framing that includes creating new approaches to intractable problems,

adapting programs to changing conditions, applying effective principles to new contexts (scaling innovation), catalyzing systems change, and improvising rapid responses in crisis conditions. Social innovation unfolds in social systems that are inherently dynamic and complex, and often turbulent. The implication for social innovators is that they typically find themselves having to adapt their interventions in the face of complexity. Funders of social innovation also need to be flexible and adaptive in alignment with the dynamic and uncertain nature of social innovation in complex systems. Developmental evaluators track, document, and help interpret the nature and implications of innovations and adaptations as they unfold, both the processes and outcomes of innovation, and help extract lessons and insights to inform the ongoing adaptive innovation process. At the same time, this provides accountability for funders and supporters of social innovations and helps them understand and refine their contributions to solutions as they evolve. Social inno-

vators often find themselves dealing with problems, trying out strategies, and striving to achieve goals emerging from their engagement in the change process that they could not have been identified before that engagement, and that continue to evolve as a result of what they learn. The developmental evaluator helps identify and make sense of these emergent problems, strategies, and goals as the social innovation *develops*. The emergent/creative/adaptive interventions generated by social innovators for complex problems are significant enough to constitute *developments* not just improvements, thus the niche for *developmental* evaluation.

Traditional evaluation approaches advocate clear, specific, and measurable outcomes that are to be achieved through processes detailed in a linear logic model. Such traditional evaluation demand for upfront, preordained specificity doesn't work under conditions of high innovation, exploration, uncertainty, turbulence, and emergence. Indeed, premature specificity can do harm and generate resistance from social innovators, as indeed it has, by constraining exploration, limiting adaptation, reducing experimental options, and forcing premature adoption of a rigid model, not because such a model is appropriate but because evaluators, funders, or other stakeholders demand it in order to comply with what they understand to be good evaluation. DE emerged as a response to criticism of traditional evaluation by social innovators and their expressed need for an alternative way to engage in evaluation of their work.

DE involves evaluative thinking throughout. Judgments of merit, worth, significance, meaningfulness, innovativeness, and effectiveness (or such other criteria as are negotiated) inform ongoing adaptive innovation. Such evaluative judgments do not just come at the end of some fixed period (e.g., a 3-year grant); rather, they are ongoing and timely. Nor are evaluation conclusions reached and rendered by the evaluator independently. DE is a collaborative, interactive process.

Being utilization-focused (see Patton, Chapter 16, this volume), and because DE unfolds in complex dynamic systems where the particular meaning and significance of information may be difficult to predetermine, making sense together of emergent findings involves developmental evaluators interpreting patterns in the data *collaboratively* with social innovators, their funders, advocates, change agents, and systems change supporters. Through this empirically focused interaction, DE becomes integral to the innovative process.

HISTORY OF DE

The first article describing DE as a distinct approach was published in 1994, when the predecessor journal to the *American Journal of Evaluation*, then called *Evaluation Practice*, featured 20 evaluation theorists and practitioners speculating on the future of evaluation. I wrote about *Developmental Evaluation* (Patton, 1994) and predicted that, in an increasingly complex world, DE would become an important evaluation option.

In 2006, I had the opportunity to coach a group of Canadian evaluators on DE in a series of workshop and consulting sessions sponsored by the J. W. McConnell Family Foundation based in Montreal. The first full-day workshop ever conducted on DE was in New Zealand in March 2009, organized, sponsored, and co-facilitated with Kate McKegg. The first DE workshop in the United States was done for The Evaluators' Institute in San Francisco in 2010.

The first book-length description and explanation of DE was published in 2011 with the subtitle: *Applying Complexity Concepts to Enhance Innovation and Use*. In 2015, a book of DE exemplars was published (Patton, McKegg, & Wehipeihana, 2015). In the past decade, DE has become recognized as a distinct evaluation approach (Dickson & Saunders, 2014; FSG, 2014; Lam & Shulha, 2015; Preskill & Beer, 2012; Stol, 2018).

FIVE DE PURPOSES

In the original formulation of DE, I identified and distinguished five distinct DE purposes:

1. *Ongoing innovation*; for example, a leadership development program that regularly revises its content and process based on emergence of new issues and new ways of communicating.
2. *Scaling principles for adaptation to new locations*; for example, disseminating agro-ecological principles of a circular economy to areas dominated by industrial agriculture.
3. *Preformative evaluation*; for example, generating new ideas and designs to address problems and needs such as a program connecting Indigenous communities around the world in support of the United Nations *Declaration on the Rights of Indigenous Peoples*.
4. *Systems change*; for example, addressing homelessness as a housing systems problem as opposed to just providing services to the homeless.
5. *Humanitarian aid*; for example, DE as rapid feedback for adaptation in support of humanitarian interventions in the face of natural disasters (volcanoes erupting, tsunamis) and political unrest (civil war, persecution of subpopulations, suppression of dissent).

What Are the Essential Elements of Developmental Evaluation?

This is the *fidelity challenge*. An experienced DE practitioner recently told me, “More often than not, I find, people say they are doing DE, but they are not.”

The fidelity challenge concerns the extent to which a specific evaluation sufficiently incorporates the core characteristics of the overall approach to justify labeling that evaluation by its designated name. Just as fidelity is a central issue in efforts to

replicate effective programs in new places (Are the replications faithful to the original model on which they are based?), evaluation fidelity concerns whether an evaluator following a particular model is faithful in implementing all the core steps, elements, and processes of that model.

I’ve seen a great many evaluations labeled *utilization-focused* that provided no evidence that primary intended users had been identified and engaged to focus the evaluation on those users’ priorities. What then are the essential elements of DE? The answer is eight essential principles: (1) developmental purpose, (2) evaluation rigor, (3) utilization focus, (4) innovation niche, (5) complexity perspective, (6) systems thinking, (7) co-creation, and (8) timely feedback.

Figure 21.1 defines and highlights each principle. From my perspective, these principles must be explicitly addressed in any developmental evaluation, but how and the extent to which they are addressed depends on situation and context. The principles serve the role of sensitizing concepts. This is a significant departure from the usual approach to “fidelity,” which has traditionally meant implementing an approach operationally in exactly the same way each time. Fidelity has meant adherence to a recipe or highly prescriptive set of steps and procedures. The principles of DE, in contrast, involve sensitizing elements that must be interpreted and applied contextually—but *must be applied in some way and to some extent if the evaluation is to be considered genuinely and fully developmental*. This means that, when I read a DE report, or talk with those involved in a DE, or listen to a DE presentation at a conference, I should be able to see/detect/understand how these essential principles of DE informed what was done and what resulted.

DE as a Distinct Evaluation Approach: The Theory–Practice Link

Because DE claims a specific purpose and niche, a common question is how DE dif-

1. **Developmental purpose principle:** Illuminate, inform, and support what is being developed, identifying the nature and patterns of development (innovation, adaptation, systems change), and the implications and consequences of those patterns.
2. **Evaluation rigor principle:** Ask probing evaluation questions, think and engage evaluatively, question assumptions, apply evaluation logic, use appropriate methods, and stay empirically grounded; that is, rigorously gather, interpret, and report data.
3. **Utilization focus principle:** Focus on intended use by intended users from beginning to end, facilitating the evaluation process to ensure utility and actual use.
4. **Innovation niche principle:** Elucidate how the change processes and results being evaluated involve innovation and adaptation, the niche of developmental evaluation.
5. **Complexity perspective principle:** Understand and interpret development through the lens of complexity and conduct the evaluation accordingly. This means using complexity premises and dynamics to make sense of the problems being addressed, guide innovation, adaptation, and systems change strategies, interpret what is developed, adapt the evaluation design as needed, and analyze emergent findings.
6. **Systems thinking principle:** Think systemically throughout, being attentive to interrelationships, perspectives, boundaries, and other the key aspects of the social system and context within which the innovation is being developed and the evaluation is being conducted.
7. **Co-creation principle:** The innovation and evaluation develop together—interwoven, interdependent, iterative, and co-created—such that the developmental evaluation becomes part of the change process.
8. **Timely feedback principle:** Time feedback to inform ongoing adaptation as needs, findings, and insights emerge rather than only at predetermined times (e.g., quarterly, or midterm and end of project).

fers from formative evaluation. DE offers an alternative to formative and summative evaluation, the classic distinctions that have dominated evaluation for five decades. In the original conceptualization, a formative evaluation served to prepare a program for summative evaluation by identifying and correcting implementation problems, making adjustments based on feedback, providing an early assessment of whether desired outcomes were being achieved (or likely to be achieved), and getting the program stabilized and standardized for summative assessment. It is not uncommon for a new program to go through 2 to 3 years of formative evaluation, working out start-up difficulties and getting the program model stabilized, before conducting a summative evaluation. Over time, formative evaluation has come to designate any evaluative efforts to *improve* a program. Improvement means making it better. In contrast, DE focuses on *adaptive development*, which means making the program different because, for example, (1) the context has changed (which comes with the territory in a complex dynamic environment); (2) the clientele have changed significantly; (3) learning leads to a significant change; or (4) a creative, innovative alternative to a persistent issue or challenge has emerged. Here are examples of DE *developments*:

- A program serving one population (White, low-income high school drop-outs) adapts to demands to serve a different population (e.g., immigrants, people coming out of prison, people with particular disabilities). This kind of adaptation goes beyond improvement. It requires developmental adaptation.
- A workshop or course moves online from the classroom. Teaching effectively online requires major adaptation of both content and process, as well as criteria for interpreting success. Again, this goes well beyond ongoing improvement.

FIGURE 21.1 DE principles.

- Public health authorities adapt to a new disease like ebola. Innovation and adaptation become the order of the day, not just improving existing procedures.

Theory Underpinnings of DE: How Systems Thinking and Complexity Theory Inform the Practice of DE

Thinking systemically is fundamental to DE. This means, at a minimum, understanding interrelationships, engaging with multiple perspectives, and reflecting deeply on the practical and ethical consequences of the boundary choices you make. The required shift in thinking is from focusing on discrete components of a program to thinking in terms of systemic relationships (Williams, 2014). Innovation often involves changing an existing system at some level and in some way. Thus, social innovators are interested in and motivated by changing systems—health systems, educational systems, food systems, criminal justice systems. In so doing, they engage in efforts and thinking that supersede traditional project and program logic models. To evaluate systems change, developmental evaluators need to be able to engage in systems thinking and treat the system (or systems) that are targeted for change as the evaluand (the thing being evaluated). This means inquiring into, tracking, documenting, and reporting on the development of interrelationships, changing boundaries, and emerging perspectives that provide windows into the processes, effects, and implications of systems change (Williams, 2005, 2008; Williams & van 't Hof, 2014). The cases in the *Developmental Evaluation Exemplars* book include changing the youth homelessness system, the early childhood system, Indigenous food systems in Africa and in the Andes, community systems where people are mired in poverty, and Ontario's school system.

These systems change cases illustrate and illuminate how a developmental evaluation is attuned to both linear and nonlin-

ear relationships, both intended and unintended interactions and outcomes, and both hypothesized and unpredicted results. Fundamental systems-oriented developmental evaluation questions include the following: In what ways and how effectively does the system function for whose interests? Why so? How are the system's boundaries perceived? With what implications? To what extent and in what ways do the boundaries, interrelationships, and perspectives affect the way the innovative change process has been conceptualized and implemented? How has social innovation changed the system, through what processes, with what results and implications?

The Complexity Perspective

Viewing innovation through the lens of complexity adds another way of framing, studying, and evaluating social innovations. Innovations involve uncertain outcomes and unfold in situations where stakeholders typically disagree about the nature of the problem and what should be done to address it (Patton, 2011, Ch. 5). In essence, complexity theory directs our attention to characteristics and dimensions of dynamic systems change, which is precisely where innovation unfolds. Core complexity theory-driven DE questions include the following: In what ways and how can the dynamics of complex systems be captured, illuminated, and understood as social innovation emerges? To what extent do the dynamics of uncertainty and disagreement shift and change during the unfolding of the innovation? How is innovation's development captured and understood, revealing new learning and knowledge that can be extrapolated or applied elsewhere?

Methods Used in DE

DE does not rely on or advocate any particular evaluation method, design, tool, or inquiry framework. A DE can include

any kind of data (quantitative, qualitative, mixed), any kind of design (e.g., naturalistic, experimental), and any kind of focus (processes, outcomes, impacts, costs, and cost-benefit, among many possibilities) depending on the nature and stage of an innovation and the priority questions that will support development of and decision making about the innovation. This can include rapid turnaround randomized controlled trials, surveys, focus groups, interviews, observations, performance data, community indicators, network analysis—whatever sheds light on key questions.

The process and quality of engagement between the primary intended users (social innovators) and the developmental evaluator is as much the method of DE as any particular design, methods, and data collection tools. Asking evaluation questions, examining and tracking the implications of adaptations, and providing timely feedback on an ongoing basis—these are the methods of DE.

Whatever methods are used or data are collected, *rapid feedback is essential*. Speed matters. Dynamic complexities do not slow down or wait for evaluators to write their reports, get them carefully edited, and then approved by higher authorities. Any method can be used but it will have to be adapted to the necessities of speed, timely reporting, and just-in-time, in-the-moment decision making. This is a major reason why the developmental evaluator should be part of the innovation team, to be present in real time as issues arise and decisions have to be made.

Methods can be emergent and flexible; designs can be dynamic. Contrary to the usual practice in evaluation of fixed designs that are implemented as planned, DE designs can change as the innovation unfolds and changes. If surveys and interviews are used, the questions may change from one administration to the next, discarding items that have revealed little of value or are no longer relevant, and adding items that address new

issues. The sample can be emergent (Patton, 2015, Ch. 5) as new participants or sites emerge and others are abandoned. Both baselines and benchmarks can be revised as new information emerges.

Developmental evaluators need to be agile, open, interactive, flexible, observant, and have a high tolerance for ambiguity. The developmental evaluator is, in part, the instrument. Because the evaluation is co-created and the developmental evaluator is part of the innovation team, bringing an evaluation perspective and evaluative thinking to the team, the evaluator's capacity to be part of the team and facilitate the evaluation elements of the innovative process involves essential people skills and is also part of the method for developmental evaluation.

In essence, how the evaluation is facilitated with meaningful involvement of primary intended users and skilled engagement of the developmental evaluator affects those users' judgments about the evaluation's credibility and utility—and their willingness to act on feedback. The active and engaged role of the developmental evaluator has been called "the art of the nudge" (Langlois, Blanchet-Cohen, & Beer, 2012, p. 39).

DE as Accountability for Learning and Adaptation

Accountability is traditionally associated with spending funds in accordance with contractual requirements to achieve set targets. But the DE approach to accountability includes *accountability for learning and adaptation*. This was the conclusion the senior staff of the Minnesota-based Blandin Foundation reached while engaged in DE focused on the Foundation's strategic framework. The result was the *Mountain of Accountability* (Blandin Foundation, 2014). DE is used to support learning, adaptation, systems change, mission fulfillment, principles-focused evaluation, and walking the talk of values. While traditional evaluations focus on improving and making decisions about

projects, DE addresses strategy implementation and effectiveness at the mission fulfillment level.

An exemplar of DE is a case study by Stol (2018) that documented how a social justice–focused teacher education program attempted to cultivate equity-driven relationships between preservice teachers and their experienced placement teachers. The DE process embodied an approach to learning that centered on the interaction between individuals catalyzed by data. The DE integrated evaluation into the fabric of the program in a way that functioned as a learning intervention for participants, while also providing meaningful data for program-level discussions and adaptations.

THE LARGER CONTEXT FOR DE

Four intersecting social change trajectories and trends converge around DE. First is the worldwide demand for *innovation*. The private, public, and nonprofit sectors are all experiencing pressure to innovate. As the world's population grows, climate change threatens, and technology innovations expand horizons and possibilities exponentially, to mention just three forces for change, social innovation is recognized as essential to address global problems. A good way to see how DE has intersected with the more general innovation trajectory over the last decade is to look at the *Stanford Social Innovation Review*, which began publishing in 2003. An archival search turns up a number of references to DE, including as “next generation evaluation” and “a game-changing approach” (FSG, 2014).

The second trajectory focuses on *systems change*. Project-level evaluation does not translate directly into systems change evaluation. Treating a *system* as a unit of analysis, that is, as the evaluand (the evaluated), requires systems understandings and systems thinking. DE brings a systems orientation to evaluating systems change.

The third trajectory flows from *complexity*. Innovation and systems thinking point to complexity theory as the relevant framework for making sense of how the world is changed.

The fourth trajectory acknowledges DE as a *legitimate evaluation approach*. I've heard from evaluators and social innovators all over the world who were already engaged in DE thinking and practices but didn't have a recognizable name for what they were doing and expressed appreciation for identifying the approach as a rigorous option.

NEW DIRECTIONS IN DE: FIVE NEW DE PURPOSES

DE is still developing—and will continue to do so. Regularly, I hear about new applications, new initiatives, and new insights being generated by developmental evaluators as reflective practitioners. Earlier in this chapter, I identified the five original purposes for and applications of DE. Here are five emergent purposes for evaluation to add to the original five.

Adaptation to Crisis

The coronavirus pandemic has revealed new purposes for DE: adapting to crisis. The coronavirus pandemic has forced programs to adapt their delivery methods, staff management practices, allocation of resources, and strategic priorities. Programs face cutbacks in staff and reduced resources to deliver programs. The necessity for these substantial and significant changes emerged quickly in March 2020. Six months later, it had become clear that the pandemic was not a temporary, short-lived challenge. It appears increasingly unlikely that there will be a return to “normal,” whatever that might mean.

The term capturing program responses to the COVID-19 pandemic is *pivot*. For example, universities had begun gradually

increasing their online offerings, but the pandemic not only accelerated online teaching but it also made it the primary, and in many cases the only, form of delivery. This meant pivoting from in-person teaching to online teaching as the dominant delivery medium in higher education. Elementary and secondary schools throughout the world have had to pivot. Public health outreach program, prisons, anti-poverty, homelessness interventions, and the full range of interventions aimed at helping people have had to be reconceptualized, adapted, and evaluated differently given the effects of the pandemic.

What once was a narrow niche for DE, namely, innovation, has become a much wider niche, namely, *adaptation to crisis*. The adaptation can take many forms: reconfiguring target populations, prioritizing services, changing the form of delivery from in-person interaction to what can be handled online, and moving from a practice-based model to a principles-focused model. DEs under crisis situations involve the evaluator in helping to develop a response to the crisis. It requires ongoing situation analysis to find out what information evaluation users need in the face of the crisis, in this case COVID-19.

Evaluators also have to pivot. An evaluation designed under the expectation that a program would have stable delivery of its intervention is quite different from one in which everything is in play for reconfiguring and adaptation. Early in the pandemic I wrote:

Adapt evaluation plans and designs now. *All evaluators must now become developmental evaluators*, capable of adapting to complex dynamics systems, preparing for the unknown, for uncertainties, turbulence, lack of control, nonlinearities, and for emergence of the unexpected. This is the current context around the world in general and this is the world in which evaluation will exist for the foreseeable future. (Patton, 2020b)

This basically means that DE is expanded

from an innovation niche to a larger, more all-encompassing *adaptation niche* that includes but is not limited to innovation.

Supporting Transformation

The second new purpose of DE is to support systems transformations. The original framing of the systems evaluation purpose in DE was to support systems change as opposed to judging the effectiveness of projects and programs. What has emerged in the context of the global climate and pandemic emergencies is a distinction between systems change and systems *transformation*. Systems change tends to involve reforming systems to make them more equitable or sustainable, incrementally. An affirmative action program might engage in institutional systems change, or developing a recycling program would involve enhancing and organizations contribution to sustainability.

In contrast, transformation involves major and dramatic development of a significantly different system than the one operating at baseline. Attacking systemic racism, or structural racism, evokes transformation in power relationships and redistribution of societal resources beyond incrementally changing regulations and policies, or better implementing appropriate policies and regulations. Incrementally improving a system to better serve those in need involves testing a theory of change. Fundamentally altering and reforming a system requires a theory of transformation. Thus, the scale, depth, and breadth of change is greater with transformation than with improvement-oriented systems change.

DE can and does serve both these purposes, but the purposes are different enough that it is worth making the distinction. The climate emergency has led to calls for transformation, as has the social justice uprising since the murder of George Floyd in June 2020. The failure of the United States to cope effectively with the coronavirus pandemic has led to calls for transforming the public health system, not just improv-

ing it. Developmental evaluators need to know whether they are in the incremental change or transformation arena, which also may mean that they help facilitate this distinction among those involved in bringing about change. Evaluating transformation means transforming evaluation *developmentally* (Patton, 2021).

Developing Networks, Alliances, and Collaborations

Efforts at systems change and transformation increasingly involve multiple actors across different organizations and programs, forming networks, alliances, and collaborations aimed at collective impact. The processes of forming, nurturing, developing, energizing, engaging, sustaining, and adapting networks of people is a distinct and increasingly critical DE purpose. The Global Alliance for the Future of Food is such a network, an alliance of some 30 philanthropic foundations working to transform food and agriculture systems based on sustainability principles. An effective, and sustainable, network or alliance must do three things: (1) provide value to participating members, (2) add value as a group acting together rather than beyond what they could accomplish individually, and (3) expand their influence through their multiple, distinct, individual, and collective networks outside the network being developmentally evaluated.

DE informs the action and alignment of these three arenas of action. They are mutually reinforcing, such that misalignment among them can undermine the effectiveness and sustainability of the network or alliance itself. The DE, then, gathers data about how members are experiencing the network, what the aggregate efforts of the network are yielding, and the ripple effects of the network beyond members. More generically, then, DE can have a distinct purpose of supporting the development of collective action as a distinct approach to systems interventions.

DE Support for Capacity Building

A fourth new direction that has emerged for developmental evaluation in the last decade is building organizational and program capacity to engage in evaluation. This often involves supporting organizational leadership in efforts to embed evaluation in the work of the organization as a way of doing business. This is where DE meets organizational development (OD), with the latter highlighting the importance of ongoing OD, and DE highlighting the importance of basing that ongoing OD on evaluative thinking, processes, and use.

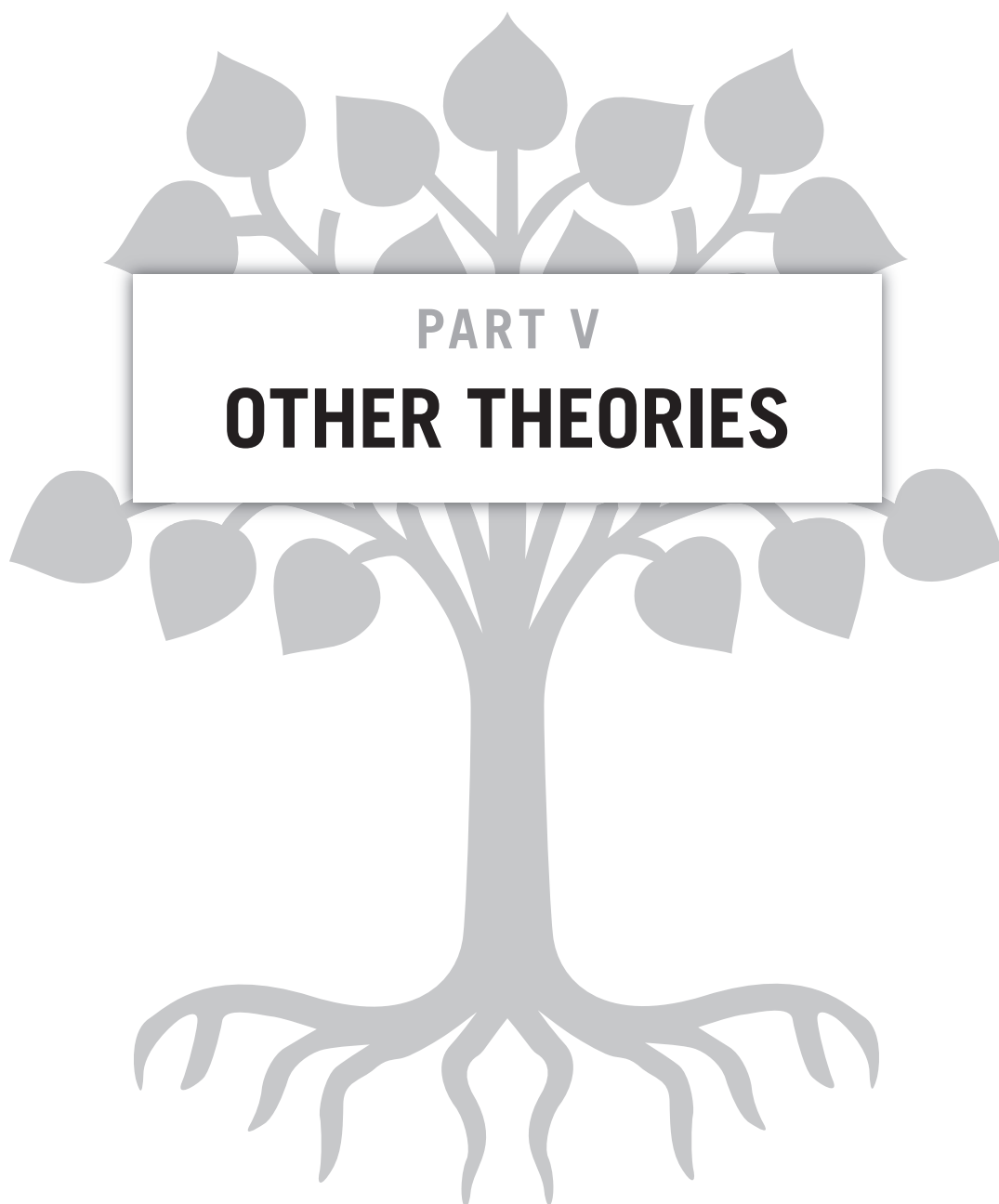
DE for Evaluation

DE, as originally conceptualized, focused on enhancing innovation and use. The focus, as has been traditional, was on supporting the development of interventions and initiatives aimed at innovation and systems change. However, in addition to those roles, and the ones described earlier, the last decade has seen the emergence of a DE function in support of the development of evaluation itself. DE as a manifestation of U-FE has given rise to *Evaluation Science* (Patton, 2018a), *Principles-Focused Evaluation* (Patton, 2018), and *Blue Marble Evaluation* (Patton, 2020a). As new approaches to evaluation emerge, as evaluation itself arises to meet those challenges, evaluation is being transformed to better evaluate transformation. New evaluation criteria for evaluating transformation have been developed (Patton, 2021). This is a developmental process in which evaluation theory and practice develop and adapt.

REFERENCES

- Blandin Foundation. (2014). Mountain of accountability. Retrieved from <http://blandinfoundation.org/resources/reports/mountain-of-accountability>.
- Dickson, R., & Saunders, M. (2014). Developmental evaluation: Lessons for evaluative practice.

- tice from the SEARCH program. *Evaluation*, 20(2), 176–194.
- FSG. (2014). *Next generation evaluation: Embracing complexity, connectivity, and change*. www.ssireview.org/nextgenevaluation
- Lam, C. Y., & Shulha, L. M. (2015). Insights on using developmental evaluation for innovating: A case study on the cocreation of an innovative program. *American Journal of Evaluation*, 36(3), 358–374.
- Langlois, M., Blanchet-Cohen, N., & Beer, T. (2012). The art of the nudge: Five practices for developmental evaluators. *Canadian Journal of Program Evaluation*, 27(2), 39–59.
- Patton, M. Q. (1994). Developmental valuation. *Evaluation Practice*, 15(4), 311–320.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.
- Patton, M. Q. (2018a). Evaluation science. *American Journal of Evaluation*, 39(2), 183–200.
- Patton, M. Q. (2018b). *Principles-focused evaluation: The GUIDE*. Guilford Press.
- Patton, M. Q. (2020a). *Blue marble evaluation*. Guilford Press.
- Patton, M. Q. (2020b, March). *Evaluation Implications of the Coronavirus Global Health Pandemic Emergency*. Blog post on Utilization-Focused Evaluation website.
- Patton, M. Q. (2021). Evaluation criteria for evaluating transformation. *American Journal of Evaluation*, 42(1), 53–89.
- Patton, M., McKegg, K., & Wehipeihana, N. (Eds.). (2015). *Developmental evaluation exemplars: Principles and practices*. Guilford Press.
- Preskill, H., & Beer, T. (2012). *Evaluating social innovation*. www.fsg.org/tabid/191/articleid/708/default.aspx?srpush=true
- Stol, T. (2018). *Cultivating equity-driven teaching partnerships: A case study of developmental evaluation in an urban teacher residency program*. Doctoral dissertation, UCLA. <https://escholarship.org/uc/item/0g02h37w>
- Williams, B. (2005). Systems and systems thinking. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 405–412). Sage.
- Williams, B. (2008). Systemic inquiry. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (Vol. 2, pp. 854–859). Sage.
- Williams, B. (2014, November 16). A systems practitioner's journey [Blog]. <https://us-mg204.mail.yahoo.com/neo/launch?partner=sbc&.rand=3sva69d24c9fe#>
- Williams, B., & van 't Hof, S. (2014). *Wicked solutions: A systems approach to complex problems*. www.bobwilliams.co.nz/ewexternalfiles/wicked.pdf



PART V

OTHER THEORIES

CHAPTER 22

Evaluation Theory Synopses

Alana R. Kinarsky
Emi Fujita-Conrads

INTRODUCTION

This edition of *Evaluation Roots* moves us closer to a discussion of *theories* of evaluation; thus, it is not restricted to the thinking and writings of individual theorists. Specifically, in this edition, we have shifted to classifying *theories* on the theory tree rather than the theorists themselves. We know, however, that some theories are quite bound to the thinking of those who have spent years, and in many cases decades, writing about a particular approach to evaluation. As such, here we discuss the work of specific theorists in relationship to specific theories because of their significant contributions to the body of work.

Understanding the ways in which a theory can be and is intended to be used in practice is a primary goal of this book. With the new and explicit emphasis on understanding evaluation theories as guides for practice, we asked well-known theorists to author chapters on how specific theoretical

approaches would be used in practice. What do we do, then, when it comes to accomplishing this goal for a theory that is closely associated with a theory writer who is no longer alive or actively engaged in writing in evaluation? This chapter addresses this dilemma.

Here, we present brief summaries of theories for which there was no theory writer to author a chapter for the book. We used the same guiding questions to write each practice summary below. Because we are not the primary writers of these theoretical approaches, our summaries are intentionally concise and they draw heavily from the work of the theory writers whose contributions significantly shaped the conceptual development of each approach. In crafting the summaries, we searched for writings and examples describing these theories used in practice. We provide references after each summary so that readers can conveniently access these writings and examples for further reading.

METHODS BRANCH

Objectives-Based Evaluation: Ralph Tyler

Originally developed to evaluate education curricula, objectives-based evaluation is commonly used to gauge the extent to which intended changes in student behaviors occur, in order to determine whether a curriculum is producing the desired outcomes (Madaus, 2011). A core belief of this approach is that educators can get students closer to their desired educational objectives by periodically evaluating their behavior, comparing it to intended effects, and revising their teaching practices based on findings (Tyler, 1942). Given its focus on using evaluation for ongoing program improvement, objectives-based evaluation is aligned with formative evaluation. Ralph Tyler was a leading proponent of objectives-based evaluation; one of his most notable contributions was the development of the National Assessment of Educational Progress (NAEP), which continues to lead educational testing and assessment in the United States (Madaus, 2011).

The first step in objectives-based evaluation is to define the educational objectives that the program or school intends to achieve. Tyler (1934) classified these objectives into eight types: (1) recalling important facts and principles; (2) understanding technical terminology; (3) drawing inferences from facts and proposing hypotheses; (4) devising ways of testing hypotheses; (5) applying the standards of technical performance, including the ability to evaluate the relative importance of several standards; (6) reporting accurate observations; (7) developing skills in particular subjects or tools, such as laboratory skills, language skills, and so on; and (8) expressing ideas effectively. Each objective should include a description of student behaviors that would be exhibited if the objective were met. Behaviors associated with a given objective can vary depending on how the program or stakeholder defines the objective. An example of an associated behavior might be a stu-

dent's ability to state and remember specific facts about a particular topic.

Next, the evaluator identifies where students are expected to display the specified behaviors. In other words, if students are asked to meet a particular educational objective, the instructor should place them in a situation where they are given a chance to show this ability. Tyler called these the "test situations" (Tyler, 1934, p. 11). How the behaviors will be measured is an important consideration when preparing for test situations, and the evaluator is encouraged to think beyond paper-and-pencil tests. As Tyler (1934) stated, "Human behavior is too complex to be measured by a single instrument along a single dimension" (p. 499). Direct observation, student work, interviews, or case study approaches are favored. Additionally, the evaluator is encouraged to use multiple criteria and methods to measure each particular outcome or aspect of achievement.

Central to this approach is teacher and stakeholder participation in determining educational objectives, behaviors, and evaluative measures that are relevant to the goals of their program or institution. In contrast to "off-the-shelf" assessments, the evaluator should develop measures specific to the educational program being studied (Madaus, 2011). Eliciting teachers' participation can help create evaluations that are representative of their programs. Furthermore, teacher involvement in the evaluation process may assist in clarifying teachers' objectives (Tyler, 1942).

When interpreting collected data, the evaluator should compare intended objectives with actual student performance to determine whether the objectives have been met. During this process, educational objectives, behaviors, and evaluation measures may need to be redefined to more accurately capture intended student outcomes. In this case, student performance can be reexamined using the newly defined measures. If there are gaps between intended objectives and actual student performance,

teachers may modify a component of their curriculum or intervention and reevaluate the intended outcome. In all, this evaluation process should lead to the continued refinement of educational objectives, revision of the curriculum or intervention, and monitoring of participant outcomes.

Experimental Methods: Donald Campbell

Experimental designs are a prized method of evaluating educational interventions due to their ability to control for bias, which can strengthen causal inferences. By employing random assignment, experimental methods enhance a study's internal validity—that is, the level of certainty that the treatment caused a change in the proposed outcomes (Campbell & Stanley, 1966). This approach is especially effective when seeking to answer summative evaluation questions for large-scale programs or policies. Donald Campbell provided some of the most significant contributions to our knowledge of experimental designs, though he never conducted an evaluation. Campbell's work on experimental methods was originally intended more broadly for social science researchers. Edward Suchman (1967) brought Campbell's theories to prominence in the field of evaluation.

The defining characteristic of a true experiment is the random assignment of the treatment condition across study units or participants. Randomization allows groups to be “probabilistically similar to each other on average,” thus ruling out possible threats that could influence causal inference before the treatment is administered (Campbell & Stanley, 1966, p. 13). A control group is included to determine how participants would have fared if no treatment had been implemented. Once the treatment and control groups are defined, study units can be assigned at random.

Following randomization, an intervention is introduced to the treatment group. The researcher then compares the treatment and control groups to determine

whether the manipulated change led to the anticipated outcomes. The evaluator should include multiple indicators of program success, viewing each as “partially imperfect and partially relevant” measures of the intended intervention outcomes (Campbell, 1991, p. 141). Additionally, the evaluator may elicit input from a range of program stakeholders, as different groups will prefer different outcomes measures.

In theory, an “experimental approach to social reform” would help alleviate societal problems by only promoting programs that withstand rigorous experimental evaluations (Campbell, 1969, p. 409). A critical consequence of using a biased evaluation is misinformed program and policy decisions. Campbell advocated for policymakers to develop “bold innovations” that could more easily be tested with experiments instead of “modest variants on established practices” (Campbell, 1991, p. 130). If these innovations are established as successful through an evaluation, then they can be scaled. Experimentalist evaluators seek to provide policymakers with dependable data on program outcomes to form conclusions (Shadish, Cook, & Leviton, 1991).

True experiments are not always feasible in real-world scenarios, however. Often, it is neither practical nor ethical to randomly assign individuals to treatment or control conditions, for example. Furthermore, these studies can be cost-intensive and require more time than is often available. For an experimentalist, however, the strength of causal inferences in determining policy and program effectiveness is worth the costs and challenges (Shadish et al., 1991).

Quasi-Experimentation: Thomas Cook

Quasi-experimentation in evaluation is aimed at strengthening causal inferences when randomization is not possible. These methods were developed in response to the difficulty of implementing true experiments in educational settings due to ethical and feasibility issues. In contrast to

experimentation, quasi-experimentation examines causal questions by comparing naturally formed groups on intervention outcome measures. The intervention is not randomly assigned; therefore, treatment and control groups may differ in systematic ways that affect their involvement in the intervention and its anticipated outcomes. These possible “rival hypotheses” could explain any observed differences between the groups (Cook & Campbell, 1986, p. 144). However, there are techniques that a quasi-experimentalist can employ to address potential threats to internal validity and strengthen causal inferences. Highly influenced by Campbell and Stanley’s (1966) work on experimentation and Cronbach’s (1982) work on enhancing study generalizations, Thomas Cook (2013) developed an approach to quasi-experimentation that seeks to maximize causal generalizations.

Careful consideration of program contextual factors in determining evaluation questions, sampling techniques, and approaches to data analysis are at the heart of the quasi-experimentation approach (Cook, 2013). To ensure that the evaluator understands the local context, Cook has urged engagement with program stakeholders in developing evaluation questions. This leads to questions that are relevant to their information needs (Cook & Gruder, 1978). The evaluator should then identify questions that can be operationalized within a given program context and that are best answered using experimental designs.

A top priority within quasi-experimentation is how participants are sampled, emphasizing methods that strengthen a study’s external validity. Although random selection is preferable for forming generalizations, it is not always feasible in educational settings. To address this, the evaluator can use “quasi-sampling” techniques when intentionally sampling for generalization (Cook, 1993). In this approach, the evaluator should purposefully choose participants who match the demographic characteristics of the populations or sites of interest.

Additionally, the evaluator should include enough participants within subpopulations of interest to allow for subgroup analyses. Subgroup analyses can be used to determine for whom the intervention is effective or if there are certain thresholds that the treatment holds. When using quasi-sampling methods rather than random selection, the evaluator must develop a strong understanding of who is in the sample and the relative effectiveness of the treatment in order to understand the conditions and settings to which the intervention can be generalized (Cook, 1993).

To strengthen internal validity, the evaluator should probe for potential alternative explanations for the observed relationship between treatment and outcome in an attempt to falsify such claims. Past research and evaluation studies as well as social science theory can be used to determine possible alternative explanations or mediating processes that could lead to the observed intervention effects (Cook & Campbell, 1986). Similarly, program theory can be used to identify mediating factors that may influence any observed outcomes. In all, Cook has argued that “all quasi-experimental data analysis should be exploratory so as to probe causal hypotheses in ways that incorporate different plausible assumptions about any alternatives that cannot be directly examined” (Cook & Campbell, 1986, p. 177).

Another option in this evaluative approach is meta-evaluation, which can enhance causal generalizations (Cook & Gruder, 1978). Single evaluation studies rarely have a large or diverse enough data pool to form broad generalizations. Meta-evaluation can be a review of extant evaluation reports or studies on a program, an empirical reevaluation of a program, or an independent replication of an evaluation study (Cook & Gruder, 1978). This approach is exemplified in Cook and colleagues’ (1975) meta-analysis of evaluations of the *Sesame Street* television program. Cook and his team used previous evaluation studies to answer new questions, including whether

the programming contributed to widening equity gaps in student achievement.

Evaluation for Improving Social Systems: Lee J. Cronbach

Evaluation for improving social systems is centered around strengthening a study's external validity (Greene, 2013) and generating knowledge about a program that can be generalized to other contexts (Shadish et al., 1991). Lee Cronbach developed three acronyms—*utos*, *UTOS*, and **UTOS*—to provide a structure for planning an evaluation for improving social systems (Greene, 2013). The lowercase *utos* refers to the sample of data collected and includes the following domains of evaluation: units (sample), treatment, observations (data collected), and setting. The uppercase *UTOS* refers to those domains within the study's population. The third acronym, **UTOS*, represents the units, treatments, observations, and settings that are distinctly different from the study's population. The goal of this evaluative approach is to understand how a study's *utos* can be used to draw inferences at the **UTOS* level. An evaluator's ability to extrapolate findings beyond the program of study is essential to informing broader educational policy decisions, including regarding future social programming.

The first step in this approach is developing the evaluation questions that will be most generalizable to other times and contexts. According to Cronbach (1982), when developing evaluation questions for generalization, the evaluator should "anticipate as best they can about the U*, T*, and O* that someone will want to know about" (p. 100). Evaluation questions should prioritize the information needs of the policy-shaping community (PSC), which may include policy and program decision makers, program administrators and participants, and external stakeholders, such as scholars or journalists (Cronbach & Associates, 1980). Accordingly, an evaluation is more useful when designed with a wider audience in mind.

To meet the needs of multiple stakeholder groups, the evaluator should prioritize increasing a study's "bandwidth" to understand a broader range of issues, over its "fidelity" to understand a smaller number of issues with more certainty (Shadish et al., 1991). Furthermore, the evaluator should determine questions for which the answer is unknown and that would help decision makers understand program processes. As Shadish and colleagues (1991, p. 338) commented, for Cronbach, "evaluation should help design better future programs rather than record past achievements."

Once the evaluation questions are determined, there are several strategies that the evaluator can employ to strengthen the generalizability of study findings. For one, generalizations from *utos* to *UTOS* can be enhanced in a study's sampling method, preferably through the use of random sampling. Second, the evaluator should explain the processes by which particular programmatic outcomes were achieved. The evaluator can form these explanations by providing in-depth descriptions of the units, treatments, observations, and settings involved in the evaluation. Qualitative data collection, such as through the use of thick description or narratives, can be useful in describing the program context, in explaining why particular quantitative outcomes were achieved, and in proposing alternative ways of achieving those outcomes.

These descriptions can facilitate the transfer of knowledge to the **UTOS* level, or to other programs or sites beyond the study's population. For instance, a decision maker may look at information about the program processes within sites that are similar in order to determine whether the intervention or program would work well for their own context (Cronbach, 1982). Along the same lines, when determining outcome measures, the evaluator should use multiple measures to increase generalizability to other contexts. Additionally, the evaluator should seek to understand how the intervention may produce differing effects for sub-

groups within the sample or sub-utos. An understanding of sub-utos is important in determining which subgroups to which the findings are generalizable.

Throughout this, the evaluator should seek to meet the needs of those within the PSC. To do so, the evaluator should be prepared to rework evaluation questions and the evaluation design by eliciting feedback from the PSC during the evaluation process. Ultimately, improvements in social systems occur when members of the PSC enhance their knowledge of the program or social problem through evaluation.

Comprehensive Evaluation: Peter H. Rossi

Comprehensive evaluation is an approach that seeks to learn about program impact, as well as how those outcomes were achieved. Peter Rossi, a trained sociologist, advanced the conception of comprehensive evaluation. The goals of a comprehensive evaluation are “to add to our body of knowledge about social processes” and improve our collective understanding of social interventions that promote better societal conditions (Rossi, 2013, p. 111). In this way, evaluation can contribute to the accumulation of knowledge regarding the effectiveness (or ineffectiveness) of programs at solving social problems, which will lead to incremental societal improvements (Rossi, 2013). Inherent in this approach is the belief that social policy decisions should occur within a democratic process, and that many factors may influence policy decisions. The evaluator should provide policymakers with comprehensive information about social programs, recognizing that this research is only one input in policy decision making (Shadish et al., 1991).

Evaluators taking a comprehensive evaluation approach should develop questions related to the conceptualization, implementation, and assessment of programs. The first step is to conceptualize the program through a theory of action—that is, a model that represents how program inputs and

program outputs are linked. Once a model is developed, the evaluator can begin to explore the extent to which the program is aligned with relevant social science theories or existing evaluations and research studies pertaining to the topic of the program. Doing so will assist the evaluator in determining whether the program has adequately conceptualized the social problem (Rossi, 2013; Rossi & Wright, 1984).

The evaluator derives evaluation questions concerning program implementation and outcomes from components of the theoretical model. In determining evaluation questions, the evaluator should consider a diversity of interests, including that of the policymaker, funder, and program (Rossi & Freeman, 1985). Implementation questions should determine whether the program is reaching the intended target population and whether it is being implemented as intended (Rossi, Lipsey, & Freeman, 2004). Last, and most important, comprehensive evaluations should determine whether the program or intervention has resulted in the desired outcomes.

In deciding on a methodological approach to use, the evaluator should determine which evaluation designs will be viewed as credible to evaluators, researchers, and policymakers (Rossi, 2013). Although qualitative methods can be used to examine program implementation, the comprehensive evaluator is concerned with their cost and feasibility. Even qualitative methods should be standardized and quantifiable—for example, through the use of structured questionnaires (Shadish et al., 1991).

When determining program impact, the evaluator is encouraged to use experimental designs. However, it is acknowledged that experimental studies can only be carried out under limited circumstances due to ethicality, time constraints, and costs (Rossi & Wright, 1984). Thus, the evaluator should choose the best evaluation design possible given the constraints involved. As Rossi noted, “The art of evaluation research may be appropriately described as an effort

to make do with considerably less than one ideally would desire” (Rossi & Wright, 1977, p. 13).

Once the impact of a program has been determined, the evaluator should calculate its cost efficacy. Specifically, even if a program produces favorable results, the overall benefits should offset the costs. Policymakers may explore alternative approaches if the program is not found to be cost-effective. These cost-efficacy determinations can take the form of cost-benefit or cost-effectiveness analyses. Overall, following this evaluation approach will lead to a comprehensive understanding of the program’s conceptualization, implementation, and impact.

Evidence-Based Policy Use: Carol Weiss

Evidence-based policy use is concerned with the use of rigorous research methods to produce evaluations that inform evidence-based policy decisions. The evaluator is encouraged to adapt his or her design and methodology to the particular context of the evaluation, taking into consideration what is being evaluated and how findings will be used:

Research quality is of value not only because of adherence to the norms of science but perhaps more importantly because it increases the power of research as ammunition in intra-organizational argument. To change minds and mobilize support, it is valuable to have research evidence of sufficient merit to withstand methodological criticism and convince others of one’s case. (Weiss, 1991, p. 195)

Rigorous and context-sensitive evaluation is a priority in evidence-based policy use as championed by Carol Weiss.

To be successful, evaluators need to understand the political climate in which they work. Given limited resources and competing priorities, the questions of which programs are funded, which are evaluated, and how they are reformed are political. Weiss (2013) summarized the evaluator’s

role in this process: “Evaluation tends to have an ‘establishment’ orientation but with a reformist slant” (p. 134). In other words, while the evaluator has a responsibility to the funders and decision makers (the “establishment”), they should also look beyond their assigned scope to understand the experiences of a broad range of stakeholders. While balancing these distinct priorities can be challenging, it is nonetheless necessary to generate innovative ideas that have the potential to improve programs and inform “fundamental changes in policy” (Weiss, 1991, p. 187).

To successfully navigate the political environment and influence decision making, evaluations must be rigorous. Evidence-based policy use is concerned with research quality but less with advocating for particular methods. As Weiss (2013) noted, “Evaluation can be qualitative or quantitative and often is both” (p. 131). Methodology should adapt to and be informed by what is being evaluated and how findings will be used. An evaluator employing the evidence-based policy use approach should be mindful of how an absence of methodological rigor can affect whether and how an evaluation is ultimately used.

A single evaluation is likely not to result in major programmatic or policy decisions. As evaluative research builds up, however, an interest in making changes tends to also increase (Weiss, 1980). Therefore, an evaluator using the evidence-based policy use approach should generate evidence that is credible and has longevity. Given this emphasis on research quality, experimental and quasi-experimental research methods are favored because of their ability to establish cause-and-effect relationships.

Another important aspect of evidence-based policy use is purposeful and inclusive stakeholder engagement. Weiss (1998, p. 30) explained: “The best way that we know to date of encouraging use of evaluation is through involving potential users in defining the study and helping to interpret results, and through reporting results

to them regularly while the study is in progress.”

Stakeholder engagement shouldn't just happen at the start or end of an evaluation; rather, it should be prioritized throughout. At the start of an evaluation, the evaluator engages all relevant stakeholders, including program staff and program recipients, to understand what information they are lacking and how they want to use information generated by the evaluation (Shadish et al., 1991). Then, throughout the evaluation, the evaluator should remain communicative with all stakeholders, candidly sharing the process and emergent findings, and seeking insights to help equalize power and encourage use. Giving all stakeholders access to the same information not only encourages evaluation use but also aids in equalizing power.

When working with stakeholders, it is also crucial to “understand the decision-making systems in organizations and the policymaking system in government if we want our evaluations to have any influence” (Weiss, 1988, p. 6). Using the Ideology, Interests, Information (III) framework, the evaluator can consider how stakeholders' ideologies, knowledge, and interests may affect the evaluation and its ultimate use (Weiss, 1983). How stakeholders align (or do not align) on interests and ideology as assessed by this framework can help the evaluator gauge if and how evaluation findings may be implemented.

Finally, the evaluator should tailor the evaluation findings to the particular context and, when applicable, to particular stakeholders. To do so, the evaluator needs to understand the context in which they work. Weiss (1991, p. 204) stated, “Researchers need to be sophisticated about the shape and contour of policy issues if their work is to be relevant to current debates.” Understanding how the field constructs and uses knowledge should inform how recommendations are drafted so that they are more likely to be reviewed and understood by stakeholders. Doing so is likely to increase

stakeholders' use of evaluation findings to make informed decisions over time, and ultimately support programs and policies that improve society.

VALUING BRANCH

Cost Analyses: Henry Levin

Cost analyses are used to predict or capture the costs of conducting a program. This information is then used to determine whether implementing the program is fiscally responsible. While cost analyses are not traditionally thought of as a key theory in evaluation, Henry Levin's work in this area has helped situate it as an important evaluative process. Levin, an economist by training, is a noteworthy education policy researcher and evaluator best known for his work on cost analyses as a means for drawing evaluative conclusions about programs.

There are three related approaches to valuing in education: cost analyses, cost-effectiveness analyses, and cost-benefit analyses (Levin, 2001; Levin & McEwan, 2001). A *cost analysis* begins with the identification of all resources necessary to implement a program or intervention. A *cost-effectiveness analysis* compares programs based on their relative costs and results. Last, a *cost-benefit analysis* is an examination of a program's cost compared to the benefit (or impact) the program is able to deliver. An essential part of all cost studies is determining how “costs” are defined, which costs are important, and how best to assess them. Because costs can be calculated and evaluated differently across strategies, it is important for the evaluator and stakeholders to clarify what they need from a cost analysis strategy and to understand how each approach works.

An evaluator can perform a cost analysis during a program's planning stage to predict what implementation will cost or during implementation to track costs. Regardless of when it is applied, a cost analysis requires a detailed accounting of all program inputs.

A list of resources, such as space, personnel, and equipment, is documented. While some of this information may be easy to track, it could also require a review of documentation, program observations, or interviews with key stakeholders.

Determining the costs of program implementation makes it possible to then conduct a cost-effectiveness or cost-benefit analysis. A cost-effectiveness analysis compares costs and outcomes across multiple programs. Doing so helps reveal which program “provides the best results of any given resource outlay or that minimizes the resource utilization for any given outcome” (Levin & McEwan, 2001, p. 2). However, identifying common outcome measures across programs is not always straightforward. For example, there could be two programs working to improve student outcomes in public high schools. One could be using test scores as the outcome measure, while the other is tracking dropout rates. In some situations, an evaluator may be able to use prior research and evaluation studies to acquire estimates of effectiveness. Given the complexity of determining common outcome measures, an evaluator engaged in a cost-effectiveness analysis should provide a detailed explanation of what data were used and how they were analyzed, so readers can properly interpret findings.

While a cost-effectiveness analysis typically compares programs, a cost-benefit analysis can focus on a single program to understand how different program models impact costs. When doing a cost-benefit analysis, the evaluator generates costs for all program inputs and also assigns a monetary value to the impact of a program. A cost-benefit analysis is a particularly strong approach when outcomes can be assessed in monetary terms and are market-oriented (Levin, 1987). Some program outcomes (e.g., reducing recidivism) are harder to monetize than others (e.g., measuring a change in earnings). Ultimately, conducting a cost-benefit analysis should reveal the program alternative that produces the

best outcomes for the lowest cost. These three approaches to cost analyses offer different views of a program’s value relative to its costs. This information can be valuable when determining which programs are most efficient at tackling societal problems.

Educational Connoisseurship: Elliot Eisner

Educational connoisseurship was developed in response to a perceived overreliance on testing with standardized measures to evaluate educational gains and outcomes. Standardized measures focus exclusively on program outcomes and, in doing so, “fail to tell the whole story” (Eisner, 1979, p. 11). In contrast, an educational connoisseurship approach uses descriptive qualitative methods to illuminate the “conditions, context, and interactions that led to these consequences” (Eisner, 1979, p. 12). Advanced by Elliot Eisner, this approach seeks to provide insights into how to improve educational practice.

Central to education connoisseurship are the notions of connoisseurship and criticism. Connoisseurship is “the ability to see, to perceive what is subtle, complex, and important” (Eisner, 1994, p. 215). In tandem, criticism is “the art of disclosing the qualities of events or objects that connoisseurship perceives” (Eisner, 1994, p. 219). A connoisseur must have an advanced knowledge of a particular subject, providing them with the ability to see qualities that another might miss. In education evaluation, this could entail a developed understanding of school processes or indicators of quality teaching (Eisner, 1979). It is this expert knowledge in education that the evaluator uses to identify successful practices.

The educational connoisseur must move their private observations to the public through the process of criticism. There are four integrated aspects of educational criticism. First is the descriptive aspect. In this step, the evaluator should attempt to describe, characterize, and portray the qualities of educational life they observe.

One intended purpose of these descriptions is to provide the reader with a vicarious experience of what the evaluator is observing. The use of film, audio recording, and photography is encouraged in order to provide a rich representation of the classroom (Eisner, 2004).

Next, the evaluator must offer an interpretation of the events observed. Applying social science theory can help make sense of the events that are occurring in the classroom. "Thick description" is commonly used to interpret these descriptive events. Third, there is an evaluation assessment in which the evaluator makes judgments about the significance of what they have observed and the value of the program. Finally, the evaluator should condense the ideas developed through the evaluation into major themes to share with educational stakeholders.

In this approach, it is the evaluator who determines what questions to ask and where to focus the evaluation study. The evaluator also forms value judgments during data collection and analysis (Eisner, 2004). Therefore, the evaluator's expertise in the subject matter is vital. They must not only be equipped with the skills necessary to guide the evaluation, but they must also have the subject-matter expertise to reveal what is important or unique about the evaluand. As a result, evaluators with different backgrounds and experiences may have different interpretations of the meaning and significance of the same intervention (Eisner, 1994).

For Eisner, this would open up a discussion of processes that are effective in education and underscore the strength in understanding a program through the various "lenses" of different educational critics (Eisner, 1994). Educational stakeholders beyond the case in question may read the evaluation reports, recognize features of their own programs, and gain insights into how they can modify their curricula or teaching (Eisner, 2004). In all, the education connoisseurship approach can facilitate an understanding of

different program processes, so that educators can improve their practice.

Fourth-Generation Evaluation: Egon Guba and Yvonna Lincoln

Fourth-generation evaluation advances an idea of evaluation as "negotiated co-creations of social reality" where the "claims, concerns and issues of stakeholders serve as organizational foci" (Guba & Lincoln, 1989, p. 50). This evaluation approach, conceived by Egon Guba and Yvonna Lincoln, prioritizes stakeholder engagement and needs while disregarding the search for an objective truth or generalizable findings. Rather, a fourth-generation evaluator seeks to understand the claims, concerns, and issues of stakeholders, people who regularly interact with the evaluand.

Fourth-generation evaluations utilize methodologies that align with a constructivist paradigm—that is, the belief that individuals "construct" their perceptions of reality. Constructive methodology involves first uncovering how stakeholders make meaning of their experience, then trying to understand how meanings differ across stakeholders. Therefore, Guba and Lincoln have urged evaluators to meaningfully engage stakeholders in determining the focus of an evaluation, data collection, and interpretation of findings (Guba & Lincoln, 1989, p. 53). In doing so, stakeholders have the opportunity to learn alongside one another, and the evaluator is less likely to disempower or disenfranchise them.

The first step of a fourth-generation evaluation is to identify the broadest scope of stakeholders, which includes anyone who has a stake in the evaluand. Once stakeholders are identified, the evaluator begins to surface their claims, concerns, and issues in relation to the evaluand. The evaluator then creates a process for sharing these findings within and across stakeholder groups to encourage understanding, discussion, and consensus building. To do this well, the

evaluator must be a skilled facilitator who is able to encourage honest discussions, navigate disagreements, and maintain respect amongst stakeholders. Additionally, the evaluator needs to be agile and respond to emerging needs and questions. When possible, the evaluator is charged with collecting and sharing additional information necessary to advance negotiations. However, Guba and Lincoln cautioned that the evaluator's "own etic (outsider) construction may be introduced for critique . . . so long as all respondents have the opportunity to criticize the evaluator's formulations as they must do with their own" (Guba & Lincoln, 1989, p. 154).

Upon completion, the evaluator drafts a case study that summarizes the process and learnings. Guba and Lincoln urged that the report contain "thick description that not only clarifies the all-important context but that makes it possible for the reader to experience it vicariously" (2001, p. 6). Multiple versions of the report may be compiled to address unique conversations or concerns of stakeholder groups. Stakeholders must be given an opportunity to discuss and negotiate the final conclusions and recommendations emerging from the evaluation.

Given fourth-generation evaluation's emphasis on meaningful stakeholder engagement, and the belief that there is no ultimate truth or generalizable findings to be uncovered, this approach to evaluation does not work well for all programs. However, programs that prioritize pluralism, meaningful engagement, and social justice may find that it is aligned with their principles and values (Huebner & Betts, 1999; Lay & Papadopoulos, 2007).

Equitable Evaluation: Asa Hilliard III

Integrating principles of equity, cultural responsiveness, and social responsibility into his evaluative practice, Asa Hilliard III made notable contributions to the field of evaluation. His experience with large-scale

evaluations, often of programs that serve marginalized youth, led him to view evaluation as a political activity that needed to meaningfully incorporate the needs and experiences of the least powerful stakeholders. He thought of evaluation as "being part of a larger process of envisioning social and educational [processes] of democratic participation" (Hood & Hopson, 2008). In his writings, Hilliard translated these goal- and situation-appropriateness considerations into several prescribed actions and stated concerns.

While Hilliard did not have a specifically defined theory, we refer to it as "equitable evaluation." Equitable evaluation approaches evaluation through situating "power and politics in contributing to the development of several fields, yet particularly education, in the era of colonization and segregation/apartheid" (Hilliard, 1984). In doing so, he did not view evaluation as neutral but rather as political and subject to power dynamics. This allows for responsiveness to the settings in which evaluations occur and consideration of their impact on those whom the results may impact. Equitable evaluation counters the tendency for evaluation to be ahistorical, acultural, acontextual, and apedagogical by recognizing the way evaluation influences what people regard as real and engaging with each of these issues that may arise in and have an impact on evaluation (Hilliard, 1989).

In his writings, Hilliard translated these goal- and situation-appropriateness considerations into several prescribed actions and stated concerns. Central to this theory of evaluation is the necessity to "reflect carefully upon present models, assumptions, and data in order to *deconstruct*, where necessary, those *structures that maintain dominance* in a society that has set democracy as its goal" (Hilliard, 1989, p. 9). Hilliard urges evaluators to focus on uplifting the "perspectives of the least powerful stakeholders in an evaluation" (Hood & Hopson, 2008). In this, the pursuit of democratic partici-

pation in evaluation requires engaging all stakeholders rather than being managed by only a few, powerful stakeholders (Hilliard, 1984). When engaging this approach Hood and Hopson comment that, “Hilliard noted that evaluation procedures were not ends in themselves but should be understood as being part of a larger process of ensuring social and education processes of democratic participation and attention to goals of equality in larger society” (2008, p. 416).

Hilliard notes that “effective evaluation requires evaluators and users to be clear about the tasks they are performing in terms of the *purpose* for evaluation, . . . the *audience* to which the results are to be directed” and in view of that, the appropriate data gathering procedure (Hilliard, 1984, p. 121). Hilliard notes that the evaluation questions to be pursued, as well as measures to be employed, must be viewed as relevant and understandable by stakeholders, including those in vulnerable populations (Hood & Hopson, 2008).

Hilliard (1985) largely focused his attention on schools and discussed categories of information outside of traditional school efforts. Many of these have broader applications to all types of evaluation. His work highlights what measurement specialists reflect as time on task. This concept applies to both the length of time stakeholders were engaged in the program and the extent of participation in program activities. He also noted concern about the match between achievement test content and the content of the curriculum. More generally, this reflected a concern that activities reflect the intended outcomes.

USE BRANCH

Context, Input, Process, and Product (CIPP): Daniel Stufflebeam

The Context, Input, Process, and Product model, commonly referred to as CIPP, is “a comprehensive framework for conduct-

ing formative and summative evaluations of projects, personnel, products, organizations, and evaluation systems” (Stufflebeam & Shinkfield, 2007, p. 325). CIPP was developed by Daniel Stufflebeam (1966) in the 1960s. Since then, it has gone through five iterations, with the latest completed in 2007 (Stufflebeam & Shinkfield, 2007). Evaluations following the CIPP model are intended to provide “ongoing evaluation services to the decision makers in an institution” (Stufflebeam & Shinkfield, 1985, p. 165). This is consistent with Stufflebeam’s belief that evaluation’s primary purpose should be to improve, not to prove.

While the CIPP model has expanded over the years, there are four primary areas of focus to which all evaluations utilizing it are likely to adhere. First, once a comprehensive contractual agreement is signed, the context for the program is explored in detail. Program beneficiaries’ needs and assets are identified to surface the needs and goals of the program. This is referred to as the *context evaluation*, and it provides stakeholders with information on intended beneficiaries, as well as more clarity on their needs. Next, an *input evaluation* examines how well a program’s strategy, work plan, and budget are responding to the assessed needs. Pertinent literature is reviewed by the evaluator, and comparable programs are identified to surface alternative strategies for consideration. Once that is complete, a *process evaluation* begins; this includes monitoring, documenting, and assessing program implementation. During this stage, the evaluation is focused on understanding how the program is being enacted. Fourth is the *product evaluation*, which assesses how well the program services have met the needs of the target beneficiaries. The evaluator is also encouraged to document whether any services were provided inappropriately or to unintended beneficiaries. While these four evaluation types are core to the CIPP model, all four need not be present, nor must they follow a particular order. Stufflebeam (2007)

encouraged evaluators to adapt to the particular program and circumstance.

A major focus of the CIPP model is program improvement and evaluation use, so the evaluator is encouraged to routinely communicate findings to key stakeholders. During each evaluation cycle, the evaluator is expected to routinely draft finding summaries that are then discussed with stakeholders in feedback sessions. Following these sessions, the evaluator revises the summaries, submitting a final report to stakeholders about twice a year. These reports are intended to help guide program implementation and improvement.

Results-Oriented Evaluation: Joseph S. Wholey

Results-oriented evaluation is centered on increasing evaluation use for decision making. This approach seeks to provide immediately useful information to high-level stakeholders who have a strong influence on programmatic decisions (Wholey, 2013). To facilitate use, the evaluator should work with stakeholders who have the ability to make program decisions to determine relevant evaluation questions and how findings can be used (Wholey, 1991, 2010). While this approach was developed through work with federal programs, it can be applied to programs at the state and local levels (Wholey, 1991).

Developed by Joseph Wholey, the results-oriented approach to evaluation is informed by his experiences evaluating federal programs at the U.S. Department of Health, Education, and Welfare and the Urban Institute. Although Wholey began his evaluation career as an experimentalist, he found that large-scale, experimental studies rarely led to program changes. For one thing, the politics involved in the formation and continuation of programs made it unlikely that a program would be ended because of unfavorable evaluation findings. Furthermore, experimental studies were costly and time-intensive, and program managers

rarely waited for evaluation results to make decisions (Wholey, 1991). Therefore, rather than answering summative evaluation questions, a results-oriented evaluation seeks to incrementally improve existing programs.

Wholey's sensitivity to resource limitations is exhibited in his notion of a "sequential purchase of information" about programs. In this approach, evaluation should only be conducted if its usefulness "outweighs the costs of acquiring it" (Wholey, 1991, p. 236). There are four evaluation processes that can be purchased sequentially within the results-oriented evaluation: (1) an evaluability assessment, (2) a rapid feedback evaluation, (3) performance monitoring, and (4) an intensive evaluation.

An *evaluability assessment* is used to determine whether conducting an evaluation will lead to program changes (Wholey, 2010). If program management does not appear willing to use evaluation findings, or if the program design and outcomes do not appear to be sufficiently clarified, then the program should not be evaluated. During this assessment, the evaluator should work with program management to clarify the intended program design in the form of a logic model. Included in the logic model should be the program's activities, expected outcomes, and causal linkages. Next, the evaluator should understand the program's actual activities through interviews, site visits, and observations. This step is used to determine whether the intended program outcomes are realistic given how the program is currently functioning. Last, the evaluator should work with management to determine how evaluation information might be used to make programmatic changes.

If the program outcomes are clearly defined and it appears that program management will act on evaluation findings, then the evaluator can conduct a *rapid feedback evaluation*. The goal of this step in the evaluation sequence is to provide program management with quick feedback on program effects using low-cost evaluation

approaches (Wholey, 2010). First, the evaluator should collect data on high-priority program outcomes as specified by program management through the evaluability assessment. Whenever possible, the evaluator should use existing program data, such as through records, monitoring reports, or past evaluations or audits. Then, program effects can be estimated by comparing the intended program outcomes with the data collected. In this stage, the evaluator should also develop alternative, more extensive evaluation designs that may be of interest to program management.

If the management team members find the evidence produced from the rapid feedback evaluation useful, they may choose to develop a *performance monitoring system*. This entails providing ongoing information on program outcomes, which can be used to determine whether the actual outcomes are aligned with the intended effects, and integrating needed changes into the program. Performance monitoring cannot lead to causal inferences concerning program effects, but it does provide descriptive data that can lead to improvements in program activities. Policymakers and management may desire information on why such outcomes have occurred, perhaps stimulating additional evaluation studies, such as case studies (Wholey, 2013).

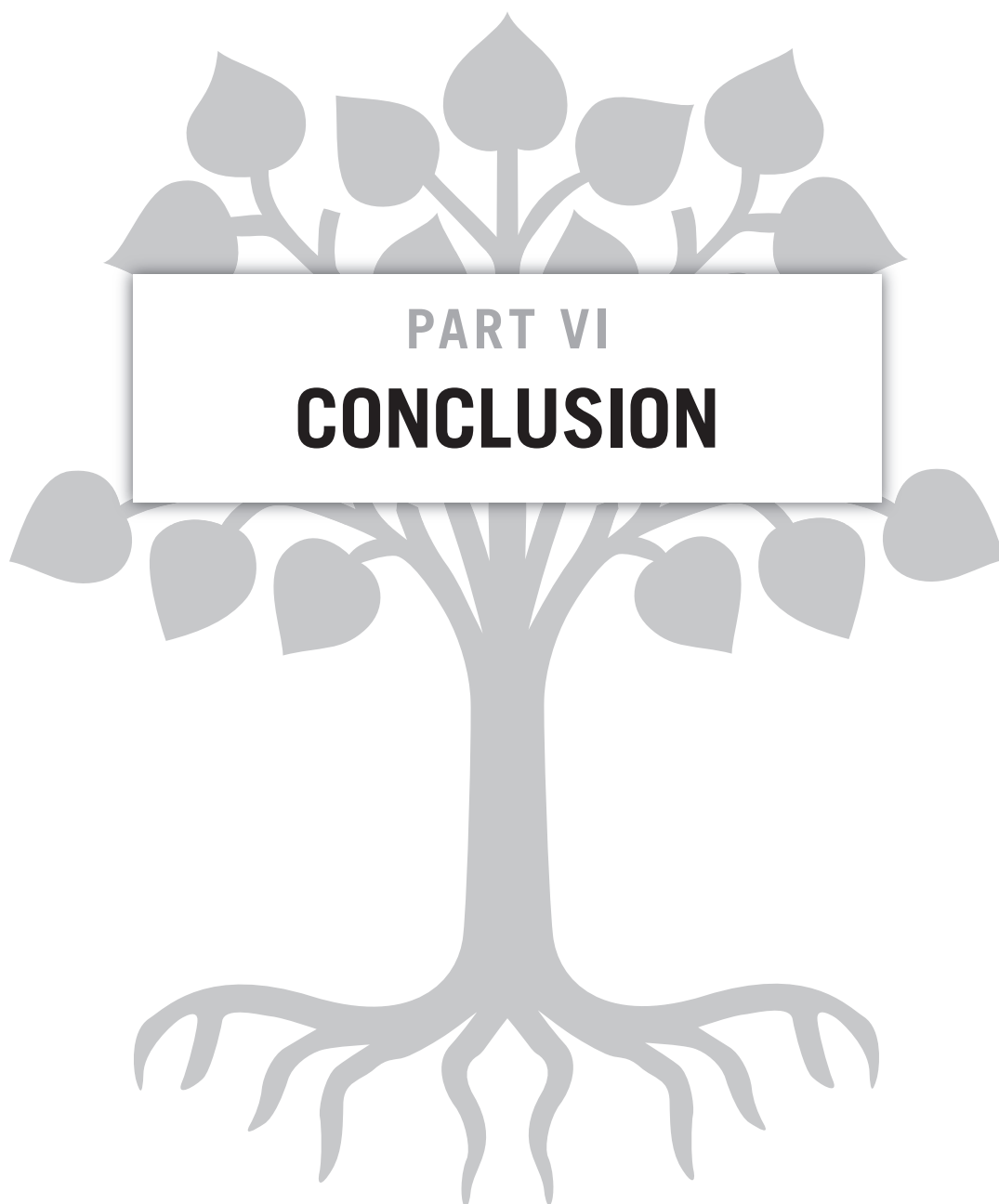
The final possible evaluation activity is an *intensive evaluation*. Intensive evaluations often take the form of experimental or quasi-experimental designs and seek to gather and analyze data that will help answer causal evaluation questions. However, Wholey found that intensive evaluations are costly and inefficient in making changes in many programs. As such, he argued that performance monitoring systems are much more cost-effective and useful to managers (Wholey, 2010). By enacting a results-oriented approach, Wholey intended to develop useful evaluations for program management decision making.

REFERENCES

- Boykin, L. (1949). Where do we go from here?: A social and educational prospectus. *Journal of Educational Sociology*, 23(2), 97–102.
- Campbell, D. T. (1969). Reforms as experiments. *Comparative Political Studies*, 24(4), 409–429.
- Campbell, D. T. (1991). Methodologist of the experimenting society. In W. Shadish, T. Cook, & L. Leviton (Eds.), *Foundations of program evaluation: Theories of practice* (pp. 119–170). SAGE.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.
- Cook, T. D. (1993). A quasi-sampling theory of the generalization of causal relationships. *New Directions for Evaluation*, 57, 39–82.
- Cook, T. D. (2013). Causal generalization: How Campbell and Cronbach influenced my theoretical thinking on this topic. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed., pp. 83–113). SAGE.
- Cook, T. D., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G., & Weber, S. J. (1975). "Sesame Street" revisited. www.russellSAGE.org/sites/default/files/sesamestreetrev.pdf
- Cook, T. D., & Campbell, D. T. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68(1), 141–180.
- Cook, T. D., & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Quarterly*, 2(1), 5–51.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. Jossey-Bass.
- Cronbach, L. J. (1991). Functional evaluation design for a world of political accommodation. In W. Shadish, T. Cook, & L. Leviton (Eds.), *Foundations of program evaluation: Theories of practice*. (pp. 323–376). SAGE.
- Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. Jossey-Bass.
- Eisner, E. W. (1979). The use of qualitative forms of evaluation for improving educational practice. *Education Evaluation and Policy Analysis*, 1(6), 11–19.
- Eisner, E. W. (1994). *The educational imagination: On the design and evaluation of educational programs* (3rd ed.). Macmillan.
- Eisner, E. W. (2004). The roots of connoisseurship and criticism: A personal journey. In M.

- C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 197–202). SAGE.
- Greene, J. C. (2013). The educative evaluator: An interpretation of Lee J. Cronbach's vision of evaluation. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed., pp. 97–105). SAGE.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. SAGE.
- Guba, E. G., & Lincoln, Y. S. (2001). *Guidelines and checklist for constructivist (aka fourth-generation) evaluation*. <https://wmich.edu/sites/default/files/attachments/u350/2018/const-eval-guba%26lincoln.pdf>
- Hilliard, A. G. (1984). Democracy in evaluation: The evolution of an art-science in context. In P. Hosford (Ed.), *What we know about teaching* (pp. 113–138). Virginia Association for Supervision and Curriculum.
- Hilliard, A. G. (1985). *Information for excellence and equity in education*. National Center for Education Statistics.
- Hilliard, A. G. (1989). Kemeti (Egyptian) historical revision: Implications for cross-cultural evaluation and research in education. *Evaluation Practice*, 10(2), 7–23.
- Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the “Nobody Knows My Name” project, and African educational excellence. *Review of Educational Research*, 78(3), 410–426.
- Huebner, A. J., & Betts, S. C. (1999). Examining fourth generation evaluation: Application to positive youth development. *Evaluation*, 5(3), 340–358.
- Lay, M., & Papadopoulos, I. (2007). An exploration of fourth generation evaluation in practice. *Evaluation*, 13(4), 495–504.
- Levin, H. M. (1987). Cost-benefit and cost-effectiveness analyses. *New Directions for Evaluation*, 34, 83–99.
- Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. *New Directions for Evaluation*, 90, 55–68.
- Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications*. SAGE.
- Madaus, G. (2011). Ralph W. Tyler's contribution to program evaluation. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (2nd ed., pp. 69–79). SAGE.
- Rossi, P. H. (1991). Comprehensive, tailored, theory-driven evaluations—a smorgasbord of options. In W. Shadish, T. Cook, & L. Leviton (Eds.), *Foundations of program evaluation: Theories of practice* (pp. 377–440). SAGE.
- Rossi, P. H. (2013). My views of evaluation and their origins. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed., pp. 106–112). SAGE.
- Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: A systematic approach* (3rd ed.). SAGE.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). SAGE.
- Rossi, P. H., & Wright, J. (1984). Evaluation research: An assessment. *Annual Review of Sociology*, 10, 331–352.
- Rossi, P. H., & Wright, S. R. (1977). Evaluation research: An assessment of theory, practice and politics. *Evaluation Quarterly*, 1(1), 5–52.
- Shadish, W., Cook, T., & Leviton, L. (Eds.). (1991). *Foundations of program evaluation: Theories of practice*. SAGE.
- Stufflebeam, D. L. (1966). A depth study of the evaluation requirement. *Theory Into Practice*, 5(3), 121–133.
- Stufflebeam, D. L. (2007). *CIPP Evaluation Model Checklist* (2nd ed.). https://wmich.edu/sites/default/files/attachments/u350/2014/cippchecklist_mar07.pdf
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). Improvement-oriented evaluation. In *Systematic evaluation: A self-instructional guide to theory and practice* (pp. 151–207). Kluwer-Nijhoff.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. Jossey-Bass.
- Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. Russell Sage Foundation.
- Tyler, R. W. (1934). *Constructing achievement tests*. Ohio State University.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35(7), 492–501.
- Weiss, C. H. (1980). Knowledge creep and decision accretion. *Knowledge*, 1(3), 381–404.
- Weiss, C. H. (1983). Ideology, interests, and information. In D. Callahan & B. Jennings (Eds.), *Ethics, the social sciences, and policy analysis* (pp. 213–235). Springer.
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, 9(1), 5–19.

- Weiss, C. H. (1991). Linking evaluation to policy research. In W. Shadish, T. Cook, & L. Leviton (Eds.), *Foundations of program evaluation: Theories of practice* (pp. 179–224). SAGE.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21–33.
- Weiss, C. H. (2013). Rooting for evaluation: Digging into beliefs. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influence* (2nd ed., pp. 130–143). SAGE.
- Wholey, J. S. (1991). Evaluation for program improvement. In W. Shadish, T. Cook, & L. Leviton (Eds.), *Foundations of program evaluation: Theories of practice* (pp. 225–269). SAGE.
- Wholey, J. S. (2010). Exploratory evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 81–99). Jossey-Bass.
- Wholey, J. S. (2013). Using evaluation to improve program performance and results. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed., pp. 261–266). SAGE.



PART VI

CONCLUSION

CHAPTER 23

Theory Influencing Practice

Marvin C. Alkin
Christina A. Christie
Naomi A. Stephen

INTRODUCTION

The subtitle of this book is *Theory Influencing Practice* and, in fact, the theory–practice relationship is very important to us. We have experienced firsthand how practice has shaped the construction of many theories. But we are less sure that theory has been properly presented so as to have the kind of impact it might have in influencing practice.

In selecting a theory, practitioners need to be assured that the theory fits their needs. They may have ideas about the kind of goals and values related to evaluation that they consider to be important. As a practitioner, what is it that you might believe are the important reasons for conducting an evaluation? What does a successful evaluation accomplish? A further consideration is context. Contexts differ, and the extent to which a theory is appropriate for a particular practitioner is dependent on the extent to which it fits his or her particular context. The evaluation activity, or potential evalu-

ation practice, may be most applicable to small- or large-scale programs, national or local programs, specific groups, and/or have a formative, summary formative, or summative purpose. Furthermore, as noted in the first chapter, the evaluation theories within this book are *prescriptive theories*; that is, certain evaluation scholars have “prescribed” how they think an evaluation should be conducted. What are the specific prescriptions, or a sample of prescriptions, that are espoused in each theory? Finally, perhaps the best descriptor of all, if you were to witness the evaluation being conducted, supposedly within the framework of a specific theory, what actions, what events or conversations, would you see that might provide evidence of the implementation of that specific evaluation theory?

In this chapter, we have displayed these elements of each theory based primarily on the chapter that theory author(s) have written. Some authors were so inspired by theoretical ideas fomenting in their current

thinking that they went well beyond a pure description of the theory for which they are known. We have attempted to develop the theory prototype based on our knowledge of these theories. In all cases, we have asked theory authors to edit our prototypes within this chapter.

We suggest that you read through these theory prototypes and consider the various elements presented and the extent to which each may or may not fit your particular value system and context in which you evaluate. Why not select several theories that seem to you most pertinent to your particular situation; then read through the description of these authors' theories as we have written them in Chapter 3. Reread the relevant chapters written by these theory authors and contemplate how the content of the particular chapter might help you to think about and conduct evaluations. If you feel that none of these theories satisfies your needs, you may want to consider examining the theories indicated in Chapter 22 for which there are not author chapters.

Furthermore, a word of caution: If you do not choose to follow a theory as prescribed by the author, it is inappropriate to indicate the name of that theory as guiding your work. Specifically, if you've not satisfied the prerequisites of conducting a theory, theorist sometimes get annoyed at the misrepresentation of their work.

THEORY PROTOTYPES

Theory Name: Randomized Field Trials

Theorist: Robert F. Boruch

Goal: Randomized field trials uses statistical theory and substantive theory about how a program is supposed to work to conduct experiments to evaluate the effectiveness of programs.

Rationale: Randomized field trials provide a scientific approach that can provide mean-

ingful evidence to prove the effectiveness of a program.

Context: Randomized field trials are best used for large-scale programs and summative evaluations.

Prescriptions (partial):

- The evaluation design requires randomization that includes “identifying and using incentives, intangible (motivation to improve) and tangible (funding to cooperate)”; “prior specification and agreement among stakeholders about eligibility and exclusions”; and “designing a trial to meet ethical and cultural constraints.”
- Groups that are subject to trials must be “regarded as independent units of analysis.”
- The experimentation accounts for “‘interactions’ that characterize how one factor may have a multiplicative effect with another factor” rather than “simply an additive effect.”
- “Substantive theory of how change is supposed to be produced by an intervention often focuses on long term effects rather than a single endpoint.”

Observable Actions (sample):

- The evaluator designs a field trial that can be “replicated” and “reproduced.”
- The evaluator must meet “quality control” standards:
 - The evaluator forms either “a control group formed by random assignment” or a “comparison group formed by using nonrandom methods.”
 - The evaluator uses “reliable and valid outcomes, . . . collected in the same way for study groups and not overaligned with the intervention.”
 - The evaluator is attentive to ensuring characteristics introduced are not “correlated with an outcome and aligned with one study group and absent from the other group.”

- The evaluator only reviews outcomes of individuals “who consent before randomization.”
- The evaluator selects “intervention and control (or comparison) groups [that are] equivalent at baseline on a preintervention measure of an outcome in the analytic sample.”
- The evaluator ensures that the experimentation adheres to all ethical guidelines, including “informed consent,” and may use approaches such as “delaying treatment,” “randomization at the cutoff point/margin of a regression discontinuing,” “‘trap door’ provisions,” or “stop wedge designs.”

Theory Name: Emergent Realist Evaluation

Theorists: Melvin M. Mark and Gary T. Henry

Goal: Emergent realist evaluation (ERE) aims for evaluation that is “tailored to the specific context in which an evaluation is planned and implemented” in order to focus on “the specific pathways available in that context for improving social conditions.”

Rationale: To “move toward social betterment” by influencing understandings and actions, “evaluation rests on the application of systematic research methods” in order “to improve the accuracy of the description of programs, the people they serve, and the outcomes they produce.”

Context: ERE is appropriate in a wide variety of contexts. ERE “locates evaluation practice squarely in a broader context of human judgment and action processes,” thereby supporting more diffuse “organizational decision making” environments as well as “accommodating narrower user-focused approaches.”

Prescriptions (partial):

- “Social betterment provides evaluators with a guiding star, a term that summarizes our concern for the well-being of others and society as a whole.” Social betterment is “the ultimate, distal goal of evaluation” but it “is necessary to identify intermediate, more proximal considerations to guide evaluation planning.”
- ERE is premised on four purposes of evaluation: “overall assessment of merit and worth, when [the] evaluation [seeks] to support bottom-line judgments; program and organizational improvement, when [the] evaluation [seeks] to point to changes in a program or the organization providing it; accountability and oversight, when evaluation is designed to provide information about a program’s compliance with legislative or other mandates; and knowledge development, when the evaluation [seeks] to expand understanding.”
- “Evaluators have at their disposal a set of methodological tools (e.g., randomized experiments, case studies) from the social sciences that can correct, at least in part, the biases and limitations on information that otherwise can affect everyday sense-making capabilities.”
- “These tools can provide stakeholders with more accurate information about a program and its outcomes, thereby increasing the capacity of stakeholders to understand and undertake well-informed actions with regard to the evaluated program or policy.”
- “Method choice is not primarily about methods, but rather about the judged likelihood of a given method or combination of methods contributing to social betterment, relative to feasible alternatives, in the anticipated decision making and action context.”
- “An evaluation . . . is of little value if the findings are not shared in ways that connect with the natural sensemaking processes of relevant parties.”

Observable Actions (sample):

- The evaluator begins with a “situational analysis” that guides a judgment about “which of the four evaluation purposes should rise to the top in evaluation planning.”
- The evaluator addresses evaluation questions related to the primary evaluation purpose in the current context, which “limit[s] the range of viable methods” and may “involve trade-offs between the ideal and the possible.”
- For example, when an assessment of merit and worth is the primary evaluation purpose, the evaluator uses “attention to rigorous research design, including quasi-experimental and experimental designs,” and “to measurement issues, such as the quality and relevance of measures used to describe program outcomes.”
- More generally, the evaluator selects from among a wide “set of methodological tools” that mitigate “the biases and limitations on information that otherwise can affect everyday sensemaking capabilities.” This increases “the capacity of stakeholders to understand and undertake well-informed actions with regard to the evaluated program or policy.”
- The evaluator may use “systematic values inquiry during the evaluation process . . . to broaden and deepen the discussion of program outcomes.”
- The “evaluators and others involved in evaluation planning should not focus on maximizing the likelihood of intended use” per se but “instead should consider the risk–reward ratio of alternative approaches to evaluation,” such as whether to prioritize an option that “has a lower likelihood of use but if influential could [greatly impact intended beneficiaries’] outcomes.”
- The evaluator recognizes “different categories of consequences that evaluation may have” at any or all of three levels—“individual, interpersonal, and collec-

tive”—and involve “four categories of evaluation consequences”: “Behavioral processes” (i.e., “changes in actions”), “Cognitive and affective processes” (i.e., “shifts in thoughts and feelings”), “General influence processes” (“sets in motion other change”), and “Motivational processes” (i.e., “goals, aspirations, and responses to perceived rewards and punishments”). The evaluator may also consider a fifth category of evaluation consequences—“relational processes [where evaluators seek to modify] aspects of ongoing relationships, structures, and organizational processes.” With awareness of these different potential interim and longer-term consequences of evaluation, the evaluator plans for evaluation influence while also being open to opportunities that may emerge.

Theory Name: Theory-Driven Evaluation

Theorist: Huey T. Chen

Goal: Theory-driven evaluation (TDE) is to assess not only whether an intervention program is effective, but also how and why the program is effective. TDE requires evaluators to assist stakeholders to clarify their program theory or theory of change. The program theory, then, is used as a framework to design the evaluation. TDE is useful to serve a program’s accountability and program improvement needs.

Rationale: An intervention program is an open system. Within the system, the program is operated under two sets of assumptions: prescriptive and descriptive. The prescriptive assumptions prescribe what actions must be taken to solve a particular problem. The descriptive assumptions describe the causal mechanisms that link the intervention to possible outcomes. TDE provides approaches to assess these two sets of assumptions and insightful information for stakeholders to better understand and improve their programs.

Context: TDE is most applicable to medium and large-scale programs.

Furthermore, TDE is useful for assessing research-oriented programs or real-world programs. Research-oriented programs usually focus on internal validity issues. TDE includes causal mechanisms in the assessment. Real-world programs must address issues on both internal and external validity. TDE is useful for including contextual factors in the assessment to provide information on program improvement and transferability.

Prescriptions (partial):

- The evaluators assist stakeholders in clarifying the action model (prescriptive assumptions) and/or the change model (descriptive assumptions) underlying their program.
- The evaluator employs a “change model [that] describes the causal process generated by the program. It makes explicit the descriptive or causative assumptions underlying causal mechanisms that connect an intervention to goals or outcomes.”
- The evaluator employs an action model, “a systematic plan for arranging staff, resources, settings, and support organizations to reach a target group and deliver intervention services. . . .” The action model consists of the following elements: implementing organization, program implementers, associated organizations/community partners, intervention and service delivery protocols, ecological context, and target group. “TD process evaluation assesses the implementation of the entire action model (intervention and contextual factors).”
- There are two theorizing approaches: “backward reasoning that starts with the change model and moves backward, . . . to obtain the arguments that justify the program” and “forward reasoning [that] considers the capacity of an organization to excel in certain areas and identifies

components that complement that capacity to ensure program success.”

- “TDE stresses that evaluators should design the evaluation to serve the needs of clients rather than prioritize a research design. Furthermore, stakeholders often need to *contextualize the intervention* precisely to fit the particularities of the community.”

Observable Actions (sample):

- The evaluator assists stakeholders in clarifying their program theory or more specifically aids in creating a sound action model and plausible change model through intensive interviews or working groups.
- The evaluator applies one of the following theorizing techniques: backward reasoning, forward reasoning, or forward/backward reasoning approaches.
- Evaluators include a change model in outcome evaluation.
- The evaluators include an action model in process evaluation.
- The evaluator conducts a multiwave formative evaluation.
- The evaluator facilitates a program stress test, “a collective brainstorming exercise that aims to identify weak structural components and functional links unlikely to withstand implementation under real-world conditions.” Evaluators facilitate the exercise to maintain the focus of the conversation and ask relevant, sometimes delicate, questions that can unearth potential weaknesses of the program.
- The application of TDE in assessing interorganizational projects provides an opportunity to understand potential strains and tensions between organizations and strategies on how to address these issues.
- “TDE proposes a realistic and practical approach, *exhibited generalization*, to address generalizability or transferability issues (Chen et al., 2014). Exhibited

generalization refers to identifying sufficient contextual factors that support the effectiveness of an intervention to inform potential stakeholders whether or not to select that intervention for their community.”

Theory Name: The Key Evaluation Checklist

Note: Unlike other summaries, this summary was primarily developed from one of Scriven’s previous works—the Key Evaluation Checklist (2013): www.michaelscriven.info/images/kec_3.22.2013.pdf.

Theorist: Michael Scriven

Goal: The Key Evaluation Checklist (KEC) strives to give an outline of ways to approach evaluations in a wide range of contexts that can be referred to continually and utilized throughout the course of an evaluation. While the checklist may be less or more comprehensive than the needs of specific evaluations, it can also be used to determine the relative worth of each aspect of the evaluation.

Rationale: The KEC offers a beneficial frame through which one is able to review evaluations to increase their efficacy and use.

Context: Proposals (or plans), projects, processes, policies, and meta-evaluation. [See General Note 1 on KEC for more specific examples of potential context for applicability.]

Prescriptions (partial):

- Program evaluation is often individualistic and the KEC encourages programs to delineate the values and needs specific to the individual program being evaluated.
- The evaluation begins with a clear and well-defined identification of the background and contexts relative to the evaluation.
- The KEC is used throughout an evaluation

and meta-evaluation (further explained in Chapter 7 of this book). It is essential to executing an effective evaluation.

Observable Actions:

- The evaluator continually thinks about the executive summary and makes adjustments throughout the process to best summarize the results.
- The evaluator clearly identifies and defines the answers to key questions from the start of the evaluation and uses this to form a written contract. These questions include who is the client and the audience, who are the stakeholders, what is the evaluator’s role, and what questions is the evaluator seeking to answer.
- The evaluator determines the impactees of the evaluation and does not use the term *beneficiaries* as it inherently assumes that all evaluation activities are beneficial.
- The evaluator conducts a strengths assessment and determines the central values driving the evaluation. The evaluator utilizes the checklist to conduct a comprehensive analysis of each.
- The evaluator conducts subevaluations that determine the worth and significance of the results of the “fact-finding” phase and the predetermined values.
- The evaluator synthesizes their findings and creates subsequent recommendations, explanations, predictions, and suggests redesigns.

Theory Name: Responsive Evaluation

Theorist: Robert E. Stake

Goal: “*Responsive evaluation* [seeks] to respond to the performance, function, situation and accomplishment of the evaluand. . . . It aims to see the value, the goodness, the shortcomings, the worth and utility of the evaluand.”

Rationale: “[As] all evaluation is shaped by the intents and outlook of the evaluators, responsive evaluators try hard to draw their findings . . . mostly from the activity and experiences within the evaluand.”

Context: Responsive evaluation is relevant to all contexts and can be “found to some extent in almost all evaluation contracts.”

Prescriptions (partial):

- “Responsive evaluators try hard to draw their findings, their probative evidence, as Michael Scriven (1994) would have it, mostly from the activity and experiences within the evaluand.”
- “Responsive evaluators do not ignore intentions and aspirations but prefer to let unstated criteria emerge from long and close acquaintance with the operation of the evaluand.”
- “Responsive evaluation is particularistic, attending to the uniqueness of the mission, context, culture, and people of the evaluand.” It is focused on this evaluand at this time, with little to no concern for comparison to other evaluands.
- “Responsive evaluators are well in touch with the staff, institution, and community, hoping to recognize the drift in priorities as time passes . . . [and] focus on the human experience of the program over a given stretch of time, with less attention to the intent at the beginning and the performance at moment’s end.”

Observable Actions (sample):

- The evaluator critically examines data and “searches for contradictory evidence. Measurement, questionnaires, and standardized interviews are used when helpful, and those results [are] critically examined.”
- The evaluator is “somewhat aware of the panorama of values, to know the better sources of relevant data, and to communicate with some who should know, such as writers and administrators. This overview

cannot be as high quality a survey as that to be done of the activity and accomplishment of the program, but knowledge of background values, the values context, needs to be realistic and pertinent.”

- “A responsive evaluation progress report or final report is likely to take the form of a story, or perhaps to include a number of stories.”

Theory Name: Deliberative Democratic Evaluation: The Evaluator as Moral Fiduciary

Theorist: Ernest R. House

Goal: In deliberative democratic evaluation, evaluators attend closely to the values and interests of stakeholders and include those in the evaluation. The goal is inclusion, discussion, and deliberation of stakeholders. In balancing and weighing these values and interests, the evaluator might act also as a moral fiduciary, protecting the interests of those less advantaged.

Rationale: American society has become increasingly fragmented, polarized, distrustful, conflicted, and unequal. These serious societal problems are manifested in programs and evaluations. Evaluators can take some responsibility for doing something about them in evaluations.

Context: Acting as a moral fiduciary is most useful in projects where there is a significant social hierarchy. This can occur in large or small, formative or summative evaluations.

Prescriptions (partial):

- Include, discuss, and deliberate with stakeholders.
- Cultivate cognitive empathy. Understand people’s predicaments as they understand them.
- Consider the social hierarchy. Race, gender, and social class are key.
- Act as a moral fiduciary.

Observable Actions:

- The evaluator carefully considers the values and interests of various stakeholders.
- The evaluator strives to understand the deep stories of major stakeholders (i.e., the interpretation of events and situations that people act on).
- The evaluator is aware of his or her own values and deep story.
- The evaluator mitigates various types of biases.
- The evaluator is transparent in his or her actions, enhancing trust.
- The evaluator interacts with stakeholders frequently, enhancing understanding.

Theory Name: Values-Engaged Democratic Evaluation

Theorist: Jennifer C. Greene

Goal: “Values-engaged democratic evaluation aspires to advance the embracement and enactment of democratic ideals and values (including ‘fairness, equity, justice, voice, as well as community and generosity, among others’) in the programs and contexts being studied.”

Rationale: Values-engaged democratic evaluation “is rooted in democratic ideals and values, especially including equity, fairness, liberty, and voice.” Evaluations are thus “designed and positioned to improve life quality, longevity, and meaningfulness for one or more targeted groups.”

Context: A values-engaged democratic approach is most appropriate for “an evaluation in which space and time are purposefully allocated to thoughtful and respectful conversations among a diversity of stakeholders, in service of greater understanding and appreciation of one another, of shared program commitments, *and* of a more democratically successful program.” This may be most appropriate and effective “for *small*

or modestly-sized program initiatives, implemented in local or regional contexts” that are “of direct importance to the local community.”

Prescriptions (partial):

- While “representatives of *all* key stakeholder groups should be included in the process of developing priority evaluation questions, constructing an evaluation design, and determining indicators of both positive and disappointing results,” “the interests and concerns of intended program participants are foregrounded as the central focus for the evaluation.”
- Key stakeholders, in rough order of priority, include “the program participants and, as relevant, their families and friends; frontline staff and allies; agency staff in charge of the program; as well as policymakers and program funders with responsibility for the program; among possible others.”
- The evaluation includes “evaluation team meetings attended by evaluators, program participants and staff, and community leaders, with active and respectful participation by many; evidence of group norms of respect, equity, and parity; and a prioritizing of needs of the program being evaluated and the people implementing this program over the needs and expectations of the evaluation and the evaluators.”
- “Effective values-engaged democratic evaluation requires time on-site, opportunities for direct interactions with program staff and participants, excellent interpersonal and listening skills, and genuine respect for the dedication and work of the program staff.”

Observable Actions (sample):

- The evaluator conducts “the following kinds of activities: regular evaluation team meetings that include representatives of all key stakeholder groups; team meetings held at places readily accessible by all key stakeholders; data collection

activities conducted at times and places convenient to program participants, and supported and facilitated by key evaluation team members; and key stakeholder names listed as contributing authors to reports on the evaluation design, implementation, and findings; among others.”

- The evaluator “invokes norms of respect, voice, equity, and caring. While such norms are often evident in the interactions and conversations of a group of people working on a collective task, actual interviews and conversations with evaluation team members could probe more deeply into character and meaningfulness of their evaluation experiences.”

Theory Name: Culturally Responsive Evaluation (CRE)

Theorist: Rodney Hopson

Goal: Culturally responsive evaluation (CRE) offers evaluation “a theoretical, conceptual, and inherently political position that affirms and amplifies the centrality of and attunement to culture in the theory and practice of evaluation.”

Rationale: CRE is rooted in a “continual interrogation of and resistance to the socializing and hegemonizing techniques of evaluation within white supremacist and racialized institutions and organizations.”

Context: CRE is used with other approaches and theories in all evaluations.

Prescriptions:

- CRE recognizes and acknowledges “the *benefits and focus of lived experiences* as important cultural context markers.”
- The practice of CRE “acknowledges and recognizes the need to dismantle hierarchies and supremacist tendencies associated most closely with White, European and European-settler knowledge traditions.”

Observable Actions:

- The evaluator does not simply engage episodically “in surface level participation and involvement in meetings, convenings, and visits.” Rather, participation is “continual and longitudinal—often intergenerational; it is not surface, but deep; and it reflects the relational and situated nature of knowledge production, attending to intersecting complexities of culture that exist in all communities, contexts, and ecologies.”
- The evaluator brings “lived experiences into the evaluation [and seeks to understand] the nature of participants’ lived experiences in ways that affirm culture.”
- The evaluator ensures that lived experiences are “integral to the development of measures, tools, resources, frameworks, and approaches that claim to be focused on those most directly and potentially negatively affected by programs, policies, and underlying situations.”

Theory Name: Transformative Evaluation

Theorist: Donna M. Mertens

Goal: “The goal of transformative evaluation is to support the development of culturally responsive interventions that foster increased respect for human rights and achievement of social, economic, and environmental justice.”

Rationale: “Evaluators have an ethical responsibility to bring to visibility and challenge versions of reality that sustain an oppressive status quo and to reveal and support those that provide a pathway toward enhanced justice and human rights.”

Context: “A transformative evaluation approach is appropriate in any context in which inequities are present that impede access for all to social, economic, and environmental justice.”

Prescriptions (partial):

- The evaluator should structure “an evaluation to be inclusive of all voices, valuing the knowledge and experiences of members of BIPOC [Black, Indigenous, and People of Color] and other marginalized communities, consciously addressing power issues, and providing safe conditions for collection of data.”
- “The transformative lens guides evaluators to conduct a contextual analysis in order to understand the dimensions of diversity that are relevant within a specific context.”
- The evaluation “design would be interrogated to ensure that the voices of all stakeholders are included, power relations are critically examined, the methods and interventions would be culturally responsive, and focus is on supporting transformative action.”
- “Transformative evaluators are quite eclectic in their choices of methodologies; however, it is likely that a mixed methods design would be chosen in order to be responsive to the full range of stakeholders’ needs.”

Observable Actions (sample):

- The evaluator “supports the development of collaborations and coalitions that are informed by the principles of social activism.”
- The evaluator incorporates “the transformative aspects of the evaluation . . . by the advocacy stance taken by
 - collaboration on behalf of an oppressed group
 - ensuring grassroots representation
 - providing a supportive environment
 - consciously addressing power issues”
- The evaluator “uses a participatory and reflective entry process into a community that is designed to build trust, address power differences, and give value to grassroots knowledge.”

- The evaluator “employs strategies throughout the evaluation that
 - build the capacity of the stakeholders to use the findings
 - disseminates findings to support increase social, economic and environmental justice.”

Theory Name: Indigenous Evaluation**Theorists:** Fiona Cram and Jill Anne Chouinard**Goal:** Culturally responsive Indigenous evaluation acknowledges the sovereign status of Indigenous peoples. It seeks to decolonize evaluation through a dual focus, namely, understanding Indigenous values, identity, knowledge and culture, and deconstructing the historical, economic, social, and political determinants of Indigenous vitality and sustainability.**Rationale:** To overcome being implicated in the ongoing colonization of Indigenous peoples culturally responsive Indigenous evaluation seeks to honor Indigenous peoples’ right to inquiry that helps fulfill their aspirations. In this way, evaluation can be in the service of Indigenous peoples and supportive of their agendas.**Context:** Culturally responsive Indigenous evaluation is appropriate when the initiatives being evaluated target or impact on Indigenous peoples, including when Indigenous peoples are a subpopulation in a larger target group. This includes initiatives delivered by Indigenous and non-Indigenous groups.**Prescriptions (partial):**

- The evaluation is shaped by the program and community context, diversity of stakeholders engaged in the process, political agendas, the evaluator’s cultural positioning (e.g., insider–outsider), the strength of relationships and community connec-

tions, institutional agenda, and dedicated time and resources.

- The foundation of culturally responsive practices within Indigenous evaluation is the evaluator's "[shared] life experience and cultural familiarity with community stakeholders, including their understanding of language and protocols."

Observable Actions (sample):

- The evaluator begins by "critically assessing his or her own worldview and versions of reality . . . [to understand] that evaluation reflects a Western understanding of reality [that] may not 'fit' with an Indigenous world."
- The "evaluator [seeks to] understand the history of the Indigenous people they work with, alongside their needs, priorities, and aspirations, and sets their evaluation within a relationship context that supports Indigenous self-determination and prosperity."
- The evaluator centers the "construction of knowledge within the context of relationships."
- The evaluator "builds collaborative relationships and engages community stakeholders in the evaluation process [to] understand context, history and cultural influences that are essential for culturally responsive practice, including the structural analysis of barriers to program success."
- The evaluator "brings the focus on the use of collaborative approaches to methodological and method choice to enhance data quality and increase the validity and trustworthiness of evaluation findings."
 - The evaluator "makes explicit the interconnections among social, political, historical, and cultural influences within the evaluation context, and asks questions about the cultural relevance and commensurability of methods and methodologies that have been developed from outside of the community,

negotiating power imbalances, especially between communities and evaluation funders."

Theory Name: Practical Policy Evaluation

Theorist: Eleanor Chelimsky

Goal: Practical policy evaluation approaches evaluation by addressing the political context in order to increase the likelihood of evaluation use for knowledge building, organizational and program change and development, or program accountability.

Rationale: Practical policy evaluation allows evaluators to "be better able to resist political pressures, survive the turbulent legislative climate, design stronger studies, and improve the likelihood of [evaluation] use" particularly given that legislative branch confrontations require that evaluations "consider, prepare for, and [are] able to answer likely objections, based on legitimate but conflicting public interest values."

Context: Evaluation of governmental policies.

Prescriptions (partial):

- Practical policy evaluation suggests "an evaluation practice that pays much closer attention to political balance, transparency, and the need to eliminate any appearance of advocacy in an environment of legitimately conflicting public interest values; and much more comprehensive inclusion of the study of the relevant past history" to the evaluation.
- Practical policy evaluations seek "recognition of an ongoing need to develop systematic credibility defenses against political or methodological attack; increase in evaluator knowledge of the relevant history and its relation to the policy context of the current proposal; attention to complex spillover issues embedded in evaluation questions; and a much more detailed

analysis of differential user perspectives in an enlarged stakeholder cohort.”

- The evaluation should “look beyond evaluator experience to the study’s historical context, in terms of three driving forces: a precipitating event; the current political balance with respect to public interest values; and longer-term issues.”
- The evaluation should “not make the problem worse [and] pay careful attention to agency sensitivities and long-term experience and knowledge, and court their willingness to experiment with reasonable modifications.”
- The evaluation should “maintain an awareness of whole-problem complexity,” “expand strategies to promote use,” and “stay flexible about the role of evaluation.”
- Eight-stage strategy for promoting use: “(1) Negotiation with all users in planning a use-focused evaluation; (2) Evaluation design to accommodate user information needs and time frame; (3) Agreement on final evaluation question(s), timing and use; (4) Performance of the evaluation; (5) Informal, separate interim briefings to congressional sponsors and agency users; (6) Review of the evaluation report by all prospective users; (7) Monitoring (by us) of reception and use of the evaluation’s findings; and (8) *Live* postmortem with users to assess lessons learned.”

Observable Actions (sample):

- The evaluator seeks to develop “credibility defenses.” These include “(1) reputational strength; (2) substantive/methodological quality; and (3) presentational effectiveness.”
 - The evaluator develops reputational strength through “a stellar advisory board, presenting both research and subject-matter expertise; publications by staff in various social scientific journals; careful attention to the staff balance of Republicans and Democrats;

networking with social science departments in various universities; the presence of staff from the relevant social science branches, possessing methodological capabilities, an ability to synthesize as well as analyze, and a bent toward innovation and risk taking; and national and international efforts to contact and learn from the experience of other evaluation units working in contexts similar to our own.”

- The evaluator encourages substantive/methodological quality through “persuasiveness in an evaluation report with respect to what has been learned, how it has been learned, and what caveats need to be imposed on that learning” through using “technical appendices” and “supplementing reports with oral briefings.”
- The evaluator obtains presentational credibility by attaining “general readability” and “concern for reasonable length, precision of logic and style, [and] the avoidance of technical language in the report (to the degree possible).”
- The evaluator strives to “(1) document—at the evaluation design stage—expected interactions between the subject area and other related systems or fields of knowledge, with special emphasis on any conflicts in policies across two agencies or areas; (2) determine objectively the likely relative importance of those interactions to the evaluation; and (3) identify any existing data sets, relating to any of those interactions, that might be available for use.”
- The evaluator stresses: “use-promoting efforts, such as frequent, regular meetings with even remote users; more accurate . . . expectations about timing and goals; willingness to research past evaluative failures and their causes; demonstrable evaluator objectivity with regard to subject matter; careful integration of

user knowledge and experience; and a much better understanding of how to target policy-action strategies for behavior change.”

Theory Name: Context-Sensitive Evaluation

Theorist: Marvin C. Alkin

Goal: “The *goal* of CSE [context-sensitive evaluation] is use,” aiding stakeholders in decision making, “establishing or altering of attitudes, or building an individual or organization’s evaluation capacity . . . [for] *this program at this time*.”

Rationale: By focusing on use and stakeholder involvement, CSE provides the greatest opportunity for the evaluation “improving *this program at this time*” as the program will be improved through small changes which will add up to have a greater effect.

Context: “The CSE approach is most applicable to small- and medium-scale programs (primarily local and not federal or large-scale state programs).”

Prescriptions (partial):

- The evaluator uses prescriptions based on factors found in evaluation use research. These factors form the four context elements in CSE.
- “Evaluator personally and intensely commits to attaining evaluation use.”
- The evaluator must be concerned about best serving the stakeholders and programs that are the recipient of their services, seeking to understand the perspectives and values of individual stakeholders, the views generally held within the program, and those at other organizational levels and uses that information to enhance the evaluation.
- The evaluation has high technical quality as appropriate to the situation leading to “a high level of evaluation use.”

Observable Actions (sample):

- “The evaluator comments on and emphasizes their commitment to the attainment of evaluation use in meetings with stakeholders. The evaluator can be seen asking questions about potential use during the evaluation.”
- “The evaluator helps stakeholders to understand use possibilities and may be seen teaching or facilitating stakeholders to engage in an analysis of use alternatives.”
- “The evaluator seeks to work with individual stakeholders/users who will (or might) have the ability to influence decisions within the program.”
- “When working with stakeholders in developing the program’s logic model, the evaluator will insert a use component following each output and each outcome in order to sensitize stakeholders to instances where use might occur.”
- “The evaluator takes a major role in ensuring the useability of the evaluation through personal dedication to use, understanding stakeholders’ perspectives, engaging stakeholders in active participation in the evaluation, finding ways to teach stakeholders about use, and providing guidance in how to use evaluation findings.”

Theory Name: Utilization-Focused Evaluation

Theorist: Michael Quinn Patton

Goal: Utilization-focused evaluation (U-FE) aims to support effective action and informed decision-making based on meaningful evidence, thoughtful interpretation, and engaged deliberation.

Rationale: “Evaluations should be judged by their utility and actual use,” and “U-FE offers an energizing and results-oriented path forward.”

Context: “U-FE is done for and with specific primary intended users for specific, intended uses [yet] can be used for any evaluation in any context” . . . “where those involved in delivering, making decisions about, and funding programs want to learn, improve, and increase effectiveness.”

Prescriptions (partial):

- “*Stay focused on use.*” “The evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way.” “For every issue that surfaces in evaluation negotiations, for every design decision, for every budget allocation, and for every choice among alternatives, keep asking, ‘How will this affect use in this situation?’”
- Evaluation questions, design, data collection, interpretation, and reporting all follow from intended use by intended users.
- U-FE requires “(1) finding and involving those who are, by inclination, information users and (2) training and incentivizing those not so inclined.”

Observable Actions (sample):

- The evaluator “develops a working relationship with intended users to help them determine what kind of evaluation they need and will use.”
- The evaluator will “be guided by the personal factor. Customize the evaluation to the participants’ knowledge, interests, motivations, and concerns.”
- The evaluator engages “through options. Deliberating on options and expressing preferences increases participants’ understanding of the implications of making certain choices and increases ownership of the decisions and recommendations that emerge from the process.”
- The evaluator observes, interprets, and adapts. “Monitor group dynamics and progress toward priority purposes to

inform pacing, changes in the process, and keeping the group moving toward desired outcomes. Adapt facilitation to what emerges.”

- The evaluator “embed[s] evaluative thinking throughout [the evaluation process].”
- The evaluator brings “before participants new directions, emergent thinking, evolving possibilities, creative opportunities, and inspiring innovations.”

Theory Name: Empowerment Evaluation

Theorist: David M. Fetterman

Goal: “Empowerment evaluation is the use of evaluation concepts, techniques, and findings to foster [program] improvement and self-determination . . . by increasing the capacity of program stakeholders to plan, implement, and evaluate their own programs.”

Rationale: “Empowerment evaluation is important because it helps people improve their program’s performance through facilitating community control [and] a sense of ownership.”

Context: “Empowerment evaluation can be used to work with small, medium, or large groups and communities” and is most commonly used for formative evaluations.

Prescriptions (partial):

- Empowerment processes are ones in which attempts to gain control, obtain needed resources, and critically understand one’s social environment are fundamental.
- The evaluation enhances self-determination and creates a sense of ownership among participants.
- Empowerment evaluation “places evaluation in the hands of community and staff members” to “make them more likely to

make decisions and take actions based on their evaluation data.”

- “Empowerment evaluation is used to help organizations learn from their experience (building on successes, learning from mistakes, and making midcourse corrections).”

Observable Actions (sample):

- The evaluator assists stakeholders in “creating a theory of action at one stage and testing it against the existing theory of use during a later stage.”
- “Specific principles to instruct practice”:
 - The evaluator “invites involvement, participation, and diversity; contributions come from all levels and walks of life.”
 - The evaluator fosters “participation and decision making [that is] open and fair.”
 - The evaluator stresses the need “to address social inequities in society.”
 - The evaluator “respects and values community knowledge.”
 - The evaluator brings to bear the knowledge of scholars to assist users in developing “evidence-based strategies” for the evaluation.
 - The evaluator “provides community members with the opportunity to collect their own data, even though it might initially be faster and easier for the evaluator to collect the same information.”
 - The evaluator “encourages that data be used to evaluate new practices, inform decision making, and implement program practices.”
- The evaluator acts as a critical friend: “(1) creating an environment conducive to dialogue and discussion; (2) providing or requesting data to inform decision making; (3) facilitating rather than leading; and (4) being inclusive and open to ideas; and (5) willing to learn.”
- The evaluator cultivates a culture of evidence, asking “for evidence or documentation at every stage, so that it becomes normal and expected to have data to support one’s opinions and views.”
- The evaluator uses a three-step approach to empowerment evaluation, helping a group “(1) establish their mission; (2) take stock of their current status; and (3) plan for the future.” [They may also use a 10-step approach, used to build capacity and cultivate self-determination.]
 - The group comes to “a consensus concerning their missions or values.”
 - “The empowerment evaluator helps members of the group generate a list of the most important activities required to accomplish organizational or programmatic goals” and “participants in the group (prioritize and then) rate how well they are doing concerning each of the activities selected, using a 1 (*low*) to 10 (*high*) scale.”
 - The empowerment evaluator facilitates a discussion about the ratings, asking for the reason and evidence for a rating. [The evidence derived from this discussion is used to guide the selection of strategies in the group’s plans for the future.]
 - The evaluator fosters discussion that plans for the future, “generating goals, strategies, and credible evidence, (to [be used in determining] whether the strategies are being implemented and whether they are effective).”
- “The [evaluator works with the group members in getting them to] make a commitment to reviewing the status of these new strategies as a group (and [indicating a willingness] to make midcourse corrections if they are not working).”
- The evaluator also helps the group members compare their baseline ratings with anticipated milestones and future ratings to monitor progress over time.

Theory Name: Practical Participatory Evaluation (through the lens of Collaborative Approaches to Evaluation)

Theorist: J. Bradley Cousins (with credit to members of the COVE research group—Lyn Shulha, Elizabeth Whitmore, Nathalie Gilbert, Hind Al Hudib).

Goal: Practical participatory evaluation (PPE) “focuses on program problem solving and improvement” through “trained evaluators working in partnership with members of the program community to co-produce evaluative knowledge” regarding “the extent to which programs are effective [and] which aspects of the program contribute to its effectiveness.”

Rationale: The rationale for PPE is predominantly *pragmatic* (“instrumental and conceptual use of findings” to supply “information for program management [and] improvement”). Depending on circumstances, the justification for PPE may draw from *political* (the process of PPE can aid in “ameliorating social inequity” and “empowering marginalized groups”), *philosophical* (“associate with our need to understand complex program phenomena at deeper levels”) and *ethical* (reflexivity, relationality, responsibility, recognition, representation reciprocity, rights—see Box 18.1) considerations.

Context: PPE is best used in “formative or summary formative” program evaluation, or “when information needs are developmental.”

Prescriptions: PPE evaluations should be approached as a “dynamic process that may change through the course of a particular project.” The evaluator’s role is “predominantly tied to his or her expertise in evaluation logic and methods [while] stakeholders participating in [PPE primarily understand] the program logic, its history, and the context within which it is operating.”

Prescriptions (partial): Action arises from the set of eight collaborative approaches to evaluation (CAE) evidence-based principles, which apply to a wide range of CAE family members.

- With stakeholders, the evaluator should develop an understanding of the justification for PPE and clarify the motivation for this particular collaborative approach as being program learning- and improvement-oriented.
- The evaluator and a limited group of primary users should “consciously develop quality working relationships” through ongoing contact.
- The evaluator and stakeholders should develop “a shared understanding of the program.”
- The evaluator should promote [a] participatory process appropriate to PPE: likely to involve balanced decision making between evaluators and primary users, and deep participation by primary users.
- The evaluator should “monitor and respond to resource availability” through “negotiating resources for the evaluation in advance [and] reevaluating during processes.”
- The evaluator should “monitor evaluation progress and quality” and protect against stakeholder self-serving bias.
- The evaluator and participating stakeholders should “promote evaluative thinking” and “inspire the active and conscious development of an organizational cultural appreciation for evaluation and its power to leverage social change.”
- The evaluator and participating stakeholders should “consider the potential for learning, capacity building, and other practical . . . consequences of the evaluation” for the realization of use.

Observable Actions: Observable actions implicating evaluators and program com-

munity members are associated with each of the eight CAE principles.

- Assess through consultation, documentation examination who are the important stakeholders and what do they value.
- Openly discuss, listen and negotiate decision-making processes and boundaries.
- Initiate exercise to develop and validate program logic model.
- Negotiate what form collaboration will take in terms of who will participate (primary users), control of decision making (balanced), and level of engagement by stakeholders (deep).
- Negotiate resources for the evaluation in advance, especially human resources; periodically reevaluate during the process.
- Take steps to ensure data quality as per professional standards of practice.
- Assess facility and comfort with systematic inquiry and develop an informal plan for growth and development.
- Plan to formally, or informally, follow through after the findings have been established, to ensure understanding and foster evidence-based decision making.

Theory Name: Learning-Centered Evaluation (Appreciative Inquiry)

Theorist: Hallie Preskill

Goal: “The goal of learning-centered evaluation theory is that evaluation is a means for ongoing learning from evaluation processes and findings, and that ultimately, evaluation contributes to decision making and action.”

Rationale: By using learning-centered evaluation, evaluations “can be a mechanism for facilitating learning from and about evaluation practice that has the potential for organization and system transformation.”

Context: “Learning-centered evaluation is a generic theory,” not specific to one context, but it is “best suited for developmental and formative evaluations” and “learning in multisite, complex geographies is likely to be more complex and challenging—but doable.”

Prescriptions (partial):

- The evaluator engages with a wide range of stakeholders and encourage evaluation capacity building in them.
- The evaluation is “conducted using collaborative, participatory, learning focused, and systems-oriented approaches that pay particular attention to racial equity and justice.”
- “Evaluations should be conducted only when there is an intention to use the findings. Such uses include instrumental, conceptual–enlightenment–knowledge, and political–symbolic–persuasive.”
- “Expected uses of evaluation processes and findings should guide an evaluation’s design” toward “questions that matter . . . reflecting the strategic information needs of the organization or community and . . . [informing] understandings, decision making, and potential actions.”
- “A learning-centered evaluation theory puts equity at the center of this work, as it strives to take an anti-racist learning stance, and [makes] a commitment to ongoing learning about equity—what it means, what it looks like, and how it is achieved, at the individual, group, organization, community, and policy levels.”

Observable Actions (sample):

- The evaluator “provides space and time for asking questions; engaging in dialogue; identifying and challenging values, beliefs, and assumptions; and reflection (both public and private).”
- The evaluator pursues “true systems change—defined as *changing the condi-*

tions that are holding the problem in place. . . . This means studying what those conditions are—policies, practices, resource flows, relationships and connects, power dynamics, and mental models, and how best to evaluate them.”

- The evaluator pursues “understanding the concepts of group dynamics, systems change and complexity theory, trust and power, organizational change and culture, adult learning, self-efficacy, racial justice, equitable evaluation, and culturally aware practices” and “facilitates meetings effectively, provides feedback, listens actively, mediates conflict, and negotiates compromise.”
- The evaluator uses an organization’s “strategy [to] inform what gets evaluated, and evaluation should inform the design, development, and implementation of strategy.”

Theory Name: Interactive Evaluation Practice

Theorist: Jean A. King

Goal: Interactive evaluation practice (IEP) seeks “to foster the constructive engagement of ‘primary intended users.’” It “has three ultimate goals”: “teach decision makers to engage collaboratively over time,” “help leaders enhance their personal evaluation skills,” and “support a more positive environment and functional working relationships.”

Rationale: IEP seeks to ensure that “its outcomes become integrated into the program’s or organization’s ongoing functioning.” There are, however, no guarantees.

Context: IEP “can be fitting for any type of evaluation,” yet “an ideal setting for this approach—although not required—includes a long-term engagement with an organization willing to commit to the extensive involvement of groups of collaborators over time.” It works well in relatively small

programs or in programs that have multiple sites that have the possibility of interacting with virtually every essential stakeholder.

Prescriptions (partial):

- Evaluators must have “(1) sincere respect for the context and people with whom [they] interact, (2) humility, which relates to awareness of [themselves] both as a person and as others perceive [them], especially in the sometimes frightening or off-putting role of evaluator, and (3) an unflinching sense of humor.”
- IEP evaluators must “commit to learning about [themselves] and how [they] fit into the big picture of the world,” assess whether “they are the right person to engage a community of group of people,” and begin the “process of decolonizing evaluation.”
- IEP evaluators should expect conflict and engage in “high-quality and transparent communication.”
- The five components for planning and managing an IEP evaluation include “(1) knowing the context well; (2) understanding two types of ‘culture’ in that context (i.e., organizational culture and that related to issues that the cultural competence/social justice literature has addressed); (3) determining who will be involved in the process, (4) organizing logistics, and (5) communicating the process and its results.”

Observable Actions (sample):

- The evaluator does not rush the evaluation process and takes time to “schedule meetings,” “conduct sessions with *all* key participants present in both body and mind,” “prepare data for interactive analysis once available,” and “frame findings and plans for next steps.”
- The evaluator connects “on a personal level” and is responsible for “clarifying diverse perspectives, stakeholder interests, and cultural assumptions” (AEA competency 3.7), “understanding and

engaging different perspectives” (AEA competency 5.2), and “facilitating shared decision making” (AEA competency 5.3).

- IEP evaluators “actively plan and implement interactions, large and small, between and among evaluation participants.”
- “IEP evaluators structure interactions that will enable participants to connect with one another for evaluative purposes.”

Theory Name: Developmental Evaluation

Theorist: Michael Q. Patton

Goal: The goal of developmental evaluation (DE) is “to inform adaptive *development* of change initiatives in complex dynamic environments . . . and support the development of innovative projects, programs, initiatives, products, organizations, and/or systems change efforts with timely feedback.”

Rationale: DE “focuses on evaluating innovations in complex dynamic environments because that is the arena in which *social innovators* are working.” As formative–summative evaluations are insufficient, DE “provides ongoing guidance to the program for adaptation and innovation with a more dynamic, real-time feedback approach.”

Context: While DE can occur in any context, “the DE niche focuses on evaluating innovations in complex dynamic environments.”

Prescriptions (partial):

- Developmental evaluators track, document, and help interpret the nature and implications of innovations and adaptations as they unfold, both the processes and outcomes of innovation, and help extract lessons and insights to inform the ongoing adaptive innovation process.

- “DE does not rely on or advocate any particular evaluation method, design, tool, or inquiry framework. . . . Asking evaluation questions, examining and tracking the implications of adaptations, and providing timely feedback on an ongoing basis—these are the methods of DE.”
- “Developmental evaluators need to be agile, open, interactive, flexible, observant, and have a high tolerance for ambiguity.”
- “DE addresses strategy implementation and effectiveness at the mission fulfillment level” rather than simply on “improving and making decisions about projects.”
- DE should build “organizational and program capacity to engage in evaluation.”

Observable Actions (sample):

- The evaluator interprets “patterns in the data *collaboratively* with social innovators, their funders, advocates, change agents, and ‘systems change supporters.’”
- The evaluator “thinks systemically. This means inquiring into, tracking, documenting, and reporting on the development of interrelationships, changing boundaries, and emerging perspectives that provide windows into the processes, effects, and implications of systems change.”
- The evaluator provides “rapid feedback for adaptation” during the evaluation process.
- The evaluator “interprets patterns in the data *collaboratively* with social innovators, their funders, advocates, change agents, and systems change supporters.”

A note on References: For references for these prototypes, please see the chapters from which the prototype was written.

CHAPTER 24

Evaluation Policy as a Bridge between Evaluation Theory and Practice The Office of Indian Health Services, a Case Example

Emi Fujita-Conrads
Christina A. Christie
Elizabeth FastHorse

INTRODUCTION

The intent of this volume is to put forward a clearer connection between evaluation theory and practice. The evaluation literature has long included arguments for the importance of evaluation theory to the field (e.g., Shadish, 1998). Evaluation theory provides evaluators with a common knowledge base and guides how we discuss, think about, and, we hope, conduct our practice (Christie & Lemire, 2019; Shadish, 1998). Yet empirical studies demonstrate that evaluation theory is not often used explicitly to inform evaluation practice (e.g., Christie, 2003). Researchers have found, however, that *evaluation policies* can mediate the influence of theory on practice (Christie & Fierro, 2012; Dillman & Christie, 2017).

Evaluation policies formalize how, when, in what ways, and with what purpose to

conduct evaluation within an organization (Christie & Lemire, 2019). We argue that there is a strong potential for evaluation policy to serve a transformative role in translating evaluation theory to evaluation practice. With this, and given the focus of this volume, we feel it is important to explore the opportunity for evaluation theory to help shape evaluation policy and, by extension, evaluation practice. To do this, we examine the evaluation policy of the Indian Health Service (IHS) and the elements of it that have clear connections to evaluation theory.

IHS: EVALUATION THEORY–POLICY ANALYSIS

The Evidence Act (2018) mandated that 24 federal departments/agencies write and implement evaluation policies. The IHS released an evaluation policy to guide

the evaluation of its funded programs in anticipation of this mandate. As part of a larger study, we examined 17 federal evaluation policies (Fierro, Kinarsky, Echeverria-Estrada, Bass, & Christie, 2022). In presenting an analysis of the policy here, we hope to show evidence of theory in policy and how the policy might indeed guide practice. Given the purpose of the IHS, we were particularly interested in understanding the extent to which culturally responsive approaches are evident in the policy.

Evaluation scholars argue for the use of culturally responsive approaches to evaluate programs serving Indigenous communities (see Cram & Chouinard, Chapter 13, this volume). These approaches position culture at the center of the evaluation to produce evidence that is aligned with a program's values, thereby increasing the credibility and usefulness of the findings to the program community. As a way to understand the program's unique cultural, historical, and political context, culturally responsive evaluators seek to build meaningful relationships and engage the community throughout the evaluation process. When working with Indigenous communities, meaningful program involvement extends beyond developing a useful evaluation. In the United States, tribal sovereignty is embedded in the U.S. Constitution, giving tribal nations the power of self-government. To respect tribal sovereignty, the program community should guide the evaluation process.

We thought that it would be instructive to examine the IHS policy to see whether aspects of culturally responsive approaches, Indigenous evaluation, and other evaluation theories are evident in the policy. It is not our intention to assess whether the policy is good or not, or to judge whether the "right" theories are evident in the policy. Our purpose is to illustrate where theory is evident in policy so that theory might be used intentionally to shape evaluation policies. To understand better the extent to which the IHS policy is grounded in culturally responsive theory, we conducted a text

analysis of the key elements of the policy. We used an empirically derived taxonomy (see Kinarsky & Christie, 2022) to structure the text analysis.¹

The IHS evaluation policy included eight components that are thought to be common to most evaluation policies:

1. *Background*: Context that helps situate the evaluation policy
2. *Goals and Purpose*: Desired outcomes of evaluations or why the agency chooses to evaluate
3. *Management*: How aspects of evaluation practice are managed
4. *Participation and Roles*: Who is involved, when, and how
5. *Design and Methods*: How evaluations are designed; methodological decisions are made
6. *Use of Evaluations*: How evaluation results are communicated and acted on
7. *Capacity Building*: Discussion of building evaluation capacity to conduct or use evaluations
8. *Ethics*: Discussion of ethical issues in evaluation

In the following section, we examine whether the evaluation policy components align with evaluation theory. By *evaluation theory*, we are not referring to the total model, but the elements found in the evaluation theories, that is, the theoretical explanations. In Chapter 23 of this volume, Alkin, Christie, and Stephen discussed the prototypes for evaluation theories, including the prescriptions and observable actions. These are the elements we will use to guide our discussion of the policy components.

Background

Policy "background" describes the organization's history of evaluation practice, motivation for developing the policy, and a definition of evaluation. While the IHS evaluation policy did not incorporate a description of the agency's history with evaluation,

it did include the agency's motivation for policy development. Specifically, the policy describes the need to meet requirements specified by statutes, regulations, and memoranda as a motivation for developing the evaluation policy. In addition, *evaluation* is defined as "the systematic collection of information about the activities, characteristics, and outcomes of programs (which may include interventions, policies, and specific projects) to make judgments about that program, improve program effectiveness, and/or inform decisions about future program development." This definition highlights the role of evaluation in (1) establishing the merit or worth of a program, (2) improving programs, and (3) informing programmatic decision making.

Goals and Purpose

The "goals and purpose" of an evaluation policy typically provide a guide for subsequent evaluation activities. The goal of the IHS policy is to establish a set of procedures to determine "planning, funding, and using information from program evaluations to assess the impact of IHS health care services, as well as functions related to the delivery of IHS health care services."

The IHS evaluation policy also affirms the agency's commitment to the use of rigorous evaluations. Specifically, "the IHS is committed to conducting and using well-designed, rigorous evaluations on a routine basis to enable programs to adhere to performance and accountability mandates, validate outcomes, and improve program effectiveness." Ultimately the policy states that the IHS will use the information obtained from evaluation to determine the "extent to which federal programs achieve intended objectives" and "to make management decisions."

Management

By *management*, we mean the general management of evaluation activities. The IHS

evaluation policy included management elements related to the resources used to support evaluation practice and the schedule for evaluation activities. The IHS policy states that the office will "ensure that sufficient evaluation capacity and resources are made available to assess program effectiveness; identify opportunities for program improvement; and inform future management decision-making." More specifically, it is recommended that 5–10% of program funding should be allocated to evaluation.

Regarding the schedule for evaluation activities, the policy states that planning activities should begin "as soon as reasonable during program planning and development" and "suggests this planning should occur as soon as the program's mission and goals are established." Additionally, the funded program should "present an evaluation plan/approach that includes evaluations across the lifecycle of the effort so that findings can be deployed for program improvement even in the early stages." In this way, the policy articulates that evaluation activities should begin at the start of program development to improve programs from the beginning. The emphasis on scheduling activities to maximize evaluation findings is aligned with use-oriented theories that seek to develop evaluations that will contribute to program decision making.

Participation and Roles

The "participation and roles" section describes how internal and external partners should be involved in the evaluation process. The policy specifies that internal, IHS organizational staff, should establish an agencywide workgroup that will "advise, support, and monitor program evaluation efforts," as well as review program evaluation plans and integrate the use of results into routine evaluation practice. Additionally, it is the role of agency staff to (1) identify how the funded effort contributes to agency priorities, (2) determine which populations are disproportionately affected by

the health issue and whether they are being targeted by the program, (3) match the evaluation design and methods, (4) develop a logic model to present program theory, (5) specify the program outcomes and measures, and (6) provide standards for reporting evaluation results.

The IHS policy outlines key actions that the external partners (i.e., grantees and external evaluators) must take to comply with the policy. The grantee should (1) agree to the program outcomes and evaluation expectations specified in the program's funding announcement, (2) describe intended activities and outcomes through a logic model, (3) develop a plan for dissemination that will "maximize program improvement for health impact," (4) provide a plan for stakeholder engagement, and (5) provide clarity on the evaluation design and data collection procedures.

The policy offers considerations of external evaluators, or those who are responsible for conducting the evaluation, and program partners. When determining the relevance of evaluation priorities, the evaluation policy states that the evaluator should consider the interests of Congress, leadership, programs, and IHS partners, including tribes and populations, served. Additionally, the policy specifies that "evaluations should be designed to address IHS's diverse programs, customers, and stakeholders; and IHS should encourage diversity among those carrying out the evaluations." The IHS policy takes the involvement of program partners a step further by stating, "IHS Offices and large programs will consult with Tribes to identify and prioritize programs for evaluation, consistent with the Tribal Consultation Policy."

There is also an emphasis on the independence and impartiality of the evaluator. As written in the policy, credibility is enhanced when the evaluation is "independent from any process involving program policy-making, management, or activity implementation." As a result, the policy specifies that "the evaluation function will be located sep-

arately from other management functions so that it is free from undue influence and so that unbiased and transparent reporting is assured."

Overall, the policy text communicates concern for being inclusive of the evaluation priorities of multiple interested groups and ensuring that tribes have a role in determining which programs should be evaluated. Practices of inclusivity are aligned with the theoretical ideas common among culturally responsive evaluation theorists, and theorists located on the valuing branch of the evaluation theory tree. However, the policy does not specify how interested groups might be involved in the evaluation beyond determining priorities.

Design and Methods

The IHS describes rigor in its evaluations as follows:

Requires ensuring that inferences about cause and effect are well founded (internal validity); requires clarity about the populations, settings, or circumstances to which results can be generalized (external validity); and requires the use of measures that accurately capture the intended information (measurement reliability and validity).

Here, *rigor* is defined in terms of internal and external validity and reliability. We see this as a divergence in practice from culturally responsive approaches, which stress that epistemology and methodology must be culturally congruent. In culturally responsive approaches, validity is enhanced when strong relationships between community partners and evaluators are developed to determine the methodological approach.

Use of the Evaluation

The IHS policy specifies how the agency should use evaluative information. First, the policy states that the IHS offices and large programs should use evaluations to help

in annual budget justifications in compliance with the Office of Management and Budget (OMB), Government Performance and Results Act (GPRA), and Government Performance Results Modernization Act (GPRMA). Consistent with its goals and purpose, the IHS evaluation policy stresses that evaluation findings should be used to “inform program improvement and accountability.”

To enhance the usability of findings, the policy finds that the office must ensure “findings are timely and relevant, to maximize their use in organization’s planning, performance reporting, budgeting, and priority-setting processes.” Additionally, the office must ensure that “evaluation findings are easily accessible to users, major constituencies, and stakeholders,” as well as develop “a process for tracking how evaluation findings are used to improve program planning, administration, implementation and oversight and outlining how evaluation findings will effect program decisions and modification.” In doing so, the agency should play a role in “coordinating and communicating evaluation activities across the IHS organizational units with overlapping or complementary missions.”

In regard to the dissemination of evaluation findings, the IHS evaluation policy specifies:

Evaluations and findings from evaluations should be broadly available and accessible, typically on the Internet. . . . Evaluation reports will present all results, including favorable, unfavorable, and null findings. The IHS will release evaluation results timely (usually within two months of a report’s completion) and will archive evaluation data for secondary use by interested researchers (e.g., public use files with appropriate data security protections).

The internal focus on using evaluations to inform program decision making is aligned with a use-oriented theoretical positioning. Although the policy seeks to promote the

use of evaluation information for program improvements and management decision making, it stops short of describing specific strategies that an evaluator might use to enhance the likelihood of evaluation use, such as how and which partners to involve in the evaluation process. In addition, similar to methods-oriented theorists, the IHS focuses on sharing findings externally, with the aim of contributing to a body of knowledge about a program or intervention.

In contrast, culturally responsive evaluators seek to develop evaluations that will be beneficial to the community. Therefore, community members should be considered vital participants throughout the evaluation process. Rather than seeking to contribute to a more extensive knowledge base about a particular program, evaluations should be helpful to the program community. Furthermore, to respect tribal sovereignty, any data shared should be done with the consent of the tribe. Although these policy statements emphasize use-oriented theories and, to a lesser extent, methods-oriented approaches, the articulation of valuing theories is minimal.

Capacity Building

While the policy did not specify external capacity-building measures, it did note that evaluation capacity might be enhanced internally at IHS. This includes integrating “program evaluation efforts into routine IHS practice” by developing “standard processes to use evaluation results to improve program development, implementation, and monitoring.” Additionally, the internal research and evaluation office is tasked with creating and maintaining an evaluation site that provides evaluation resources to IHS personnel and partners, including training materials, a repository of internal evaluation activities, and previous and existing evaluation projects, to name a few. Using a culturally responsive approach with Indigenous communities requires understanding

tribal sovereignty and advancing community self-determination. One way to support community self-determination is to build a program's capacity to conduct its own evaluations. Although the policy provides actions to enhance the agency's capacity to inform management decision making, less text pertains to building program capacity to do and use evaluations.

Ethics

With respect to ethics, the IHS evaluation policy states:

The IHS-sponsored evaluations will be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants. Evaluations will comply with both the spirit and the letter of relevant requirements such as regulations governing research involving human subjects.

In this statement, IHS broadly defines ethical evaluation practices as aligned with human subjects' regulations. Culturally responsive evaluators seek to extend foundational views of ethical practice by promoting fairness, equity, and inclusion in the evaluation. This is more critical when working with Indigenous communities because of our history with colonizing research practices, where researchers exploited tribal nations for their own gains. Therefore, an ethical evaluation involves obtaining consent from community members before the evaluation begins, determining community members' goals and expectations for the evaluation, and designing a study that will lead to community betterment.

Our purpose in this chapter is to understand the extent to which evaluation theory, particularly culturally responsive approaches, is evident in the IHS evaluation policy. Overall, our text analysis reveals that the policy does offer some alignment with culturally responsive evaluation practices, related to tribal consultation and the

relevance of evaluation priorities. In addition, we saw policy text aligned with the theoretical leanings of methods and use-oriented theories, particularly the policy's conception of "rigor," as well as its emphasis on developing evaluations to inform management decision making. In the following section, we provide a commentary on a bottom-up, culturally responsive approach to developing evaluation policy as a method to strengthen the cultural relevancy of an evaluation policy.

COMMENTARY ON DEVELOPING AN EVALUATION POLICY

A Bottom-Up, Culturally Responsive Approach

Either intentionally or unintentionally, evaluation theory is embedded in evaluation policies, thereby influencing evaluation practice (Christie & Lemire, 2019). Obviously, we believe that it is both advantageous and beneficial to *intentionally* include evaluation theory into evaluation policy. This may be the most prudent mechanism for connecting theory to practice.

A coauthor of this chapter, Elizabeth FastHorse, is a past employee of the Indian Health Council, Inc. (IHC, Inc.) on the Rincon Indian Reservation. Established in 1970, the IHC is a consortium of nine tribes whose board members review, discuss, and evaluate health care services, funding, and employment. The clinic provides culturally appropriate health care and related services, onsite programs, periodic services, and outreach services. In total, the IHC provides health care services to approximately 15,000 clients. The IHC empowers Native wellness and promotes health and wellness for not only for its clients but also future generations.

Important for the IHC is that it is a consortium that acts as an oversight committee for how evaluation is implemented and, more importantly, ensures that evaluation is used to uphold the health needs of patients

at tribal health care facilities throughout Indian country. This work follows the same “bottom-up” process the IHC uses to identify health care needs, working with their own tribes, and thus having a deep understanding of their community’s particular contexts that impact the wellness of Native wellness and promotes culturally appropriate health care. For example, the IHC expanded its services to the Santa Ysabel reservation, which sits on the slopes of the Volcano mountains, and has limited medical facilities/equipment, in addition to no access to Internet services. Therefore, the reservation could not receive the medical services needed due to its remote context, and many tribal members did not have transportation to get to the IHC on the Rincon Indian reservation.

In developing an evaluation policy, it is important to understand the community context and the tribal needs. A culturally responsive process, for example, like how one provides culturally responsive services, is also recommended when developing an evaluation policy. First, a culturally responsive policy would explicitly acknowledge and respect the tribal self-governance of tribal nations and be guided by each tribe. The policy would be adaptable to meet the specific needs of each tribal community, including remote locations. It would help to guide evaluations that are responsive to tribal community context, health care needs, and most importantly, respects the tribal community.

This helps to produce policies with integrity, and that will help to close the gap between evaluation policy and evaluation practice. It seems that principles and concepts from culturally responsive evaluation approaches would be helpful in guiding the policy development process, as well as intentionally woven into the policy itself as a guide for conducting evaluations of programs. A “bottom-up” policy development approach guided by culturally responsive evaluation theoretical principles would help to position tribal communities as *the* priority and the types of services or medical capacity needed for a particular community.

CONCLUSION

In this chapter, we have argued that evaluation policy can play a transformative role in translating evaluation theory to practice. To examine the evidence of theory in policy, we considered the salience of various evaluation theories in the IHS evaluation policy. Given the importance of using culturally responsive approaches when working with Indigenous communities, we focused on the extent to which culturally responsive theories were present in the policy. We found that, although the IHS evaluation policy did include aspects of culturally responsive evaluation theories, it did not take full advantage of the rich literature on the topic (see Cram & Chouinard, Chapter 13, this volume). Limited cultural relevance in evaluation policies could impact the credibility, useability, and trustworthiness of the evaluations produced under the guidance of those policies by the program community.

While it is the recommendation that policymakers incorporate evaluation theory into evaluation policy, attention should be paid to how and which theories are included. In this way, policymakers could design policies that meet the unique needs and cultural context of their organization. By tailoring theory to context, we argue, the policies have great potential to increase the usefulness of evaluation work in these settings, as well as promote evaluations that are viewed as credible and useful to the program community. In an effort to promote responsive evaluation policies, we offer a commentary on how evaluation policy development may follow a “bottom-up” approach. We see a culturally responsive approach to policy development as a way to encourage the conduct of evaluations that meet the needs of Indigenous communities.

NOTE

1. The Kinarsky and Christie (2022) framework was developed by testing the applicability of two theoretically constructed taxonomies

across federal and philanthropic evaluation policies (see AEA Evaluation Policy Taskforce, 2007; Trochim, 2009).

REFERENCES

- American Evaluation Association. (2007). *AEA Evaluation Policy Task Force charge*. <https://www.eval.org/Policy-Advocacy/Evaluation-Policy-Taskforce>
- Christie, C. A. (2003). What guides evaluation?: A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–36.
- Christie, C. A., & Fierro, L. A. (2012). Evaluation policy to implementation: An evaluation of scientifically based research in practice. *Studies in Educational Evaluation*, 38(2), 65–72.
- Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation*, 40(4), 490–508.
- Dillman, L. M., & Christie, C. A. (2017). Evaluation policy in nonprofit foundation: A case study exploration of the Robert Wood Johnson Foundation. *American Journal of Evaluation*, 38, 60–79.
- Fierro, L., Kinarsky, A., Echeverria-Estrada, C., Bass, N., & Christie, C. (2022). The importance of implementation: Putting evaluation policy to work. In M. M. Mark & N. R. Hart (Eds.), *Evaluation policy. New Directions for Evaluation*, 173, 49–62.
- Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 52 Stat. 8; 132 Stat. 5529-5557 (2019). <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>
- Kinarsky, A. R., & Christie, C. A. (2022). Analysis of evaluation policies in the philanthropic sector. *American Journal of Evaluation*, 43(2), 175–192.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19, 1–19.
- Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. *New Directions for Evaluation*, 123, 13–32.

Author Index

- Abma, T. A., 31, 107, 108
Acree, J., 222
Adhikari, S., 136
Aguilar, C., 139
Ai, X., 31
Al Hudib, H., 212, 215, 223
Alkin, M. C., 6, 7, 8, 9, 10, 12, 14, 15, 26, 31, 38, 40, 41, 42, 46, 74, 100, 115, 126, 161, 173, 174, 176, 180, 184, 211, 212, 214, 231, 233, 235, 236, 285, 293
Allen-Platt, C., 55, 63
Amon, J. J., 139
Anastasiou, C. J., 31
Anderson, D., 61
Anderson, G., 152
Anderson, W., 147
Anglin, K., 63
Arbesú, M. I., 108
Argyris, C., 197, 208, 232
Aristotle, 101
Arko-Achemfuor, A., 139
Astin, A., 46
Atkin, M., 9
Aton, K., 147
Ayoo, S., 237

Bailly, B., 167, 169
Bandura, A., 208
Barnhardt, C. L., 139
Barteneuk, J. M., 205, 213
Bass, N., 293
Battiste, M., 131
Baur, V., 107
Beauchamp, E., 139

Bebchuk, L., 111
Becho, L. W., 184
Beer, T., 232, 244, 248
Behrman, J. R., 58
Bellury, L. M., 82, 86
Berk, R. A., 21
Bernard, C., 55
Betts, S. C., 265
Bhaskar, R., 75
Birkeland, S., 26
Blanchet-Cohen, N., 248
Bledsoe, K. L., 87
Bloom, B. S., 19
Bloom, H. S., 60
Bolinson, C., 141
Bonilla-Silva, E., 125
Borenstein, M., 61
Borges-Méndez, R., 130
Borry, P., 101
Boruch, R. F., 20, 21, 55, 56, 57, 63, 274
Bowman, N., 237
Bowman-Waapalaneexkweew, N. R., 35
Boykin, L., 37, 147
Boyle, S., 227
Braga, A. A., 57
Brant Castellano, M., 147
Brodhead, D., 184
Brousselle, A., 184
Brown, A., 37, 147
Brown, C. H., 60
Brown, K. K., 137
Bryant, A., 205
Bucuvalas, M. J., 26, 27, 261

Caldwell, L. D., 138
Campbell, 24
Campbell, D. T., 8, 17, 18, 19, 20, 21, 23, 25, 26, 62, 75, 87, 114, 257, 258
Campbell, R., 10, 31, 74, 112, 113
Campbell-Patton, C.E., 193
Camus, A., 171
Canales, G., 141
Caracelli, C., 107
Caracelli, V. J., 72
Carden, F., 46
Carden, F., 126
Carlson, R., 225
Carter, L., 150
Casey, J., 138
Cassaro, D. A., 131
Catsambas, T., 232
Cavino, H., 127, 131
Chalas, A., 222
Chalmers, I., 60
Chambers, D. L., 21
Chan, W., 60
Chaplowe, S. G., 239
Charmaz, K., 205
Checkel, J. T., 78
Checkoway, B., 196
Chelimsky, E., 35, 39, 40, 161, 162, 164, 167, 168, 169, 170, 283
Chen, H. T., 15, 23, 24, 25, 78, 79, 82, 85, 86, 87, 88, 89, 276, 277
Chilisa, B., 131, 138, 153, 154
Chinman, M., 208
Chouinard, J. A., 35, 138, 148, 214, 216, 282, 293, 298

- Christie, C. A., 14, 17, 26, 38, 43, 115, 118, 184, 211, 212, 233, 292, 293, 297, 298
- Christie, T., 161
- Clayton M., 139
- Clinton, J., 194
- Coffman, J., 232
- Collins, P. H., 131, 152
- Connell, J. P., 78
- Conner, R. F., 236
- Conrad, J. A., 137
- Contandriopoulos, D., 184
- Contreras, G., 108
- Cook, T. D., 8, 14, 21, 22, 62, 87, 114, 162, 164, 257, 258
- Cooper, H., 60
- Cooperrider, D., 232
- Corsaro, N., 59
- Coryn, C. L. S., 10, 78
- Cousins, J. B., 42, 43, 72, 118, 148, 180, 184, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 222, 223, 239, 288
- Covert, R., 127
- Cram, F., 35, 46, 138, 147, 148, 150, 154, 216, 282, 293, 298
- Cramer, K., 113
- Crenshaw, K. W., 130
- Cronbach, L. J., 9, 20, 23, 24, 25, 26, 31, 100, 102, 108, 151, 164, 237, 258, 259
- Dagenais, C., 184
- D'Angelo, R., 232
- Davidson, J., 95
- Davis, J.E., 127
- Davis, S., 106, 127
- Dean-Coffey, J., 138, 229, 232
- Deitz, J., 194, 206
- Delaney, L., 194
- Delpit, L. D., 154
- DeMoya, D., 20
- Dennis, R. M., 125
- D'Errico, S., 139
- Desmond, M., 116
- Diaz-Puente, J.M., 194
- Dickson, R., 244
- Diesing, P., 14
- Dillman, L. M., 292
- Donaldson, C. A., 78
- Donaldson, S. I., 85, 87, 118, 208, 236
- Dong, N., 57
- Donner, A., 57
- D'Ostie-Racinea, L., 184
- Drill, K., 139
- DuBois, W. E. B., 114
- Duffy, J. I., 204
- Dugard, P., 57
- Dunst, C. J., 196
- Durkheim, É., 16
- Earl, L. M., 42, 43, 212, 231
- Easley, J. A., 31
- Echeverria-Estrada, C., 293
- Eiler, M., 198
- Eiseley, L., 100
- Eisner, E. W., 29, 30, 225, 263, 264
- Ellett, F., 6, 9
- Ellett, F. S., Jr., 8
- Ellul, J., 170, 171
- El-Saddik, K., 139
- Elteren, A., 107
- Engel, R. S., 59
- Englehart, M. D., 19
- Espino, S. L. R., 34
- Estrella, R., 136
- Evans-Lee, M., 194
- Eyong, G., 232
- Farrington, D., 59
- Fatima, S., 58
- Feagin, J. R., 114
- Fehler-Cabral, G., 112
- Fernande-Moral, M. J., 194
- Fernandez, J. S., 194
- Fetterman, D. M., 10, 41, 42, 74, 193, 194, 198, 199, 200, 204, 205, 206, 207, 215, 286
- Fierro, L. A., 292, 293
- Fierro, R. S., 239
- Fine, M., 131
- Fishman, S., 131
- Fitzpatrick, J. L., 236
- Flecha, R., 139
- Fleischer, D. N., 17
- Floden, R., 170
- Freeman, H. E., 9, 25, 260
- Freeman, M., 107
- Fried, J., 111
- Frierson, H. T., 107, 138, 147
- Frierson, H.T., 128
- Fulbright-Anderson, K., 78
- Furst, E. J., 19
- Gandhi, A. G., 26
- Gaotlhogogwe, M., 131
- Garnier, H. E., 58
- Garrow, C. E., 36
- Geertz, C., 16
- Gerstner, C.C., 55, 63
- Gesundheit, N., 194, 206
- Ghoshal, S., 205
- Gilbert, N., 212, 215, 223, 288
- Glass, G. V., 8, 9, 60
- Glennester, R., 59, 60
- Gooler, L. E., 236
- Gopal, S., 228, 232
- Gordon, E. W., 125, 132
- Greenberg, D., 75
- Greene, J. C., 9, 27, 33, 107, 118, 231, 259, 280
- Greenwood, D. J., 194
- Greenwood, S., 137
- Grodzicki, D., 184
- Gruder, C. L., 22, 258
- Guba, E. G., 9, 31, 36, 38, 107, 173, 264, 265
- Gueron, J., 58, 60
- Gutiérrez, E., 228
- Guttmann, A., 116
- Guzman, B. L., 148
- Habashi, J., 139
- Hacker, J. S., 111
- Hall, M., 8, 229
- Hamilton, D., 9
- Hanberger, 107
- Handa, S., 57, 58
- Hansen, M., 212
- Hargraves, M., 232
- Hastings, T., 9
- Hatry, H. P., 39, 118
- Hattie, J., 194
- Haugh, K., 184
- Hayes, A. F., 58
- Hedges, L. V., 60, 61
- Henry, G. T., 6, 15, 22, 23, 25, 38, 68, 69, 72, 73, 74, 214, 275
- Herbert, J. L., 75
- Heron, J., 151
- Herr, K., 152
- Hesse-Biber, S., 34
- Higgins, J. P. T., 61
- Hill, W. H., 19
- Hilliard, A. G., 37, 127, 265, 266
- Himmelfarb, G., 163
- Hirschman, A., 171
- Ho, J. E., 14
- Hobbes, T., 16
- Hochschild, A. R., 113, 114, 116
- Hoffmann, D., 139
- Hofstetter, C., 31
- Hood, S., 33, 34, 37, 87, 107, 128, 129, 130, 131, 138, 147, 229, 265, 266
- Hopson, R. K., 33, 34, 37, 107, 126, 128, 138, 145, 147, 148, 150, 229, 265, 266, 281
- House, E. R., 8, 9, 27, 31, 32, 33, 74, 102, 110, 114, 115, 116, 120, 151, 173, 225, 231, 235, 236, 279
- Howe, K. R., 27, 32, 33, 74, 116
- Huarng, K.H., 213
- Huberman, M., 8, 9
- Hudib, H., 288
- Huebner, A. J., 265
- Hughes, E., 180
- Hughes, G. B., 128, 147
- Hunt, D., 213
- Hunt, M. M., 60
- Hureau, D. M., 57
- Ikeda, J., 139
- Imm, P., 208
- Inouye, T. E., 136
- Irwin, K., 149
- Isaza, G. T., 59
- Israel, B. A., 196
- Jackson, R.E., 147
- Jamal, A., 139
- Jamison, D. F., 37
- Jewell, M. E., 163
- Johnson, K., 180

- Jorgensen, M., 36
 Julnes, G., 15, 22, 68
- Kaftarian, S., 194
 Kahneman, D., 114, 187
 Kania, J., 228
 Kawakami, A., 147, 154
 Kennedy, D. M., 110
 Kerr, S., 127
 Kho, A., 69
 Kimble, L., 82
 Kinarsky, A. R., 293, 298
 King, J. A., 7, 44, 45, 127, 184, 212, 231, 234, 235, 236, 237, 238, 239, 240, 242, 290
 Kirkhart, K. E., 8, 23, 33, 34, 35, 37, 38, 72, 130, 147, 150
 Klar, N., 57
 Kovach, M., 35, 131
 Kramer, K., 116
 Kramer, M., 228
 Krathwohl, D. R., 19
 Krugman, P., 111
 Kubisch, A. C., 78
 Kuttner, R., 171
 Kwaymullina, A., 151
- Labin, S., 204
 LaFrance, J., 8, 35, 131, 147, 229
 Lahoff, R., 139
 Lai, M., 147
 Lam, C. Y., 244
 Langhout, R. D., 194
 Langlois, M., 248
 LaPointe, N., 196
 Lavrakas, P., 58
 Lay, M., 265
 Leal, P. A., 149
 Lefebvre, H., 148
 Leget, C., 101
 Leithwood, K. A., 72, 180
 Lemire, S. T., 26, 184, 292, 297
 Lesesne, C. A., 204
 Lessig, L., 111
 Levin, H. M., 30, 262, 263
 Levin, M., 194
 Leviton, L. C., 8, 14, 62, 164, 180, 257
 Lewin, K., 194, 213
 Lewis-Chapman, H., 136
 Lian, B., 88
 Liebenberg, L., 139
 Lim, C. Y., 222
 Lincoln, Y. S., 31, 36, 107, 264, 265
 Linnan, L., 46
 Lipsey, M. W., 25, 260
 Lopez, A., 128
 Lucks, D., 139
- MacDonald, B., 102, 107, 120, 147, 225
 Mack, K., 228
 Madaus, G. F., 18, 19, 256
 Madison, A. M., 127, 131
 Magro, C., 135, 189
 Maikuri, A., 130
 Major, T. E., 131
- Malembeka, G., 139
 Mandell, M., 75
 Mark, M. M., 6, 15, 22, 23, 38, 68, 69, 71, 72, 73, 74, 75, 118, 214, 275
 Marsick, V., 232
 Marx, K., 16
 Masia, B. B., 19
 Massey, D., 147, 148, 150
 Mas-Tur, A., 213
 Mathison, S., 17, 231
 Matsumura, L. C., 58
 Maynard, R., 57, 61
 McEwan, P. J., 30, 262, 263
 McKegg, K., 244
 McManus, H. D., 59
 McSweeney, A. J., 21
 Merchant, S., 222
 Merrill, B., 139
 Mertens, D. M., 34, 35, 46, 135, 136, 137, 138, 139, 140, 141, 147, 148, 212, 215, 229, 281
 Meyers, D. C., 204
 Miles, M., 8
 Miller, F., 125
 Miller, K., 60
 Miller, R. L., 10, 74, 140, 174
 Milofsky, C., 130
 Mithaug, D. E., 196
 Mlambo-Ngcuka, P., 137
 Mohr, B., 232
 Mokgolodi, H., 131
 Montesquieu, 16
 Morosanu, L., 82, 86, 88
 Morris, M., 8
 Mosteller, F., 21, 55, 57
 Moynihan, D. P., 168
 Mulder, S. S., 135, 189
 Murphy-Graham, E., 26
 Murray, D. M., 57
- Newcomer, K. E., 39, 118
 Nichols, R., 35, 131, 147
 Nierse, C., 107
 Noakes, L. A., 10, 78
 Nobles, W., 145
 Nsangi, A., 57
- Okun, T., 128
 Olbrechts-Tyteca, L., 173
 Onstott, M., 75
 Ormond, A., 150
- Pan, H.L.W., 82
 Papachristos, A. V., 57
 Papadopoulos, I., 265
 Parlett, M., 147
 Parsons, W., 232
 Patterson, S. C., 163
 Patton, M. Q., 4, 7, 9, 41, 42, 45, 68, 71, 75, 103, 124, 131, 147, 186, 187, 188, 189, 190, 191, 197, 215, 231, 235, 239, 244, 247, 248, 250, 251, 285, 291
 Pawson, R., 78
 Peak, G. L., 128, 131
- Peck, L., 60
 Perelman, C., 173
 Persig, R., 100
 Peters, L., 131
 Petrosino, A., 26
 Pham, L., 68
 Piciotto, R., 118
 Pierson, P., 111
 Piketty, T., 111
 Piontek, M. E., 227
 Ployhart, R. E., 205, 213
 Popham, W. J., 8, 9, 225
 Porima, L., 147
 Poth, C. N., 222
 Powell, J. A., 232
 Powell-Threets, K., 88
 Prah, K., 154
 Preskill, H., 43, 44, 74, 208, 227, 228, 232, 239, 244, 289
 Price, A. W., 137
 Puigvert, L., 139
- Qureshi, C., 139
- Ramirez, R., 184
 Ramos, M., 139
 Rappaport, J., 196
 Rawls, J., 32, 116
 Ray, V., 128
 Reagan, T., 150
 Reason, P., 151
 Reichardt, C. S., 62
 Reviere, 153
 Reyes, K., 139
 Ridde, V., 184
 Riecken, H. W., 56
 RMSA Team, 217
 Robertson, P., 36
 Rodriguez, A. V., 139
 Rodríguez-Campos, L., 215
 Rog, D. J., 236
 Rogers, P. J., 78, 85
 Rogoff, B., 151
 Rollock, D., 125
 Rolston, H., 58, 60
 Romm, N., 139
 Rossi, P. H., 15, 21, 23, 24, 25, 26, 78, 260, 261
 Rothstein, B., 111
 Rothstein, H., 61
 Rousseau, J. J., 16
 Rubin, D., 62
 Ruby, A., 63
 Russ-Eft, D., 239
 Ryan, K. R., 118
 Rykaczewska, 75
 Rynes, S. L., 205
- Saez, E., 111
 Sanders, J. R., 8, 9
 Sastre-Merino, S., 194
 Saunders, M., 244
 Sayer, A., 101
 Scheurich, J. J., 125
 Schochet, P. Z., 57

- Scholer, D., 78
 Schon, D. A., 197, 208, 232
 Schröter, D. C., 10
 Schulz, A., 196
 Schwandt, T. A., 8, 110, 112, 126
 Schwartz, A., 239
 Scriven, M., 17, 20, 27, 28, 29, 30, 31, 74, 101, 103, 114, 225, 234, 236, 278, 279
 Searle, M., 222
 Sefa Dei, G. J., 146
 Seidenfeld, D., 57, 58
 Senese, G., 145
 Senge, P., 208, 228, 232
 SenGupta, S., 34, 145
 Serolong, L., 139
 Shadish, W. R., 8, 9, 14, 19, 25, 26, 27, 28, 62, 87, 114, 164, 257, 259, 260, 261, 262, 292
 Shanker, V., 127, 128, 130
 Shaw, I., 118
 Shaw, J., 112
 Shinkfield, A. J., 266
 Shulha, L. M., 180, 212, 215, 217, 218, 222, 223, 231, 244, 288
 Simons, H., 120
 Singer, J. D., 58
 Small, M. L., 112
 Smart, D. H., 239
 Smith, L. T., 35, 145, 147, 149, 152, 153
 Snow, R., 24
 Snyder, B., 20
 Soderstrom, E. J., 21
 Soja, E., 149
 Spybrook, J., 58, 61
 Stack, E. E., 139
 Stake, R., 9, 20, 27, 28, 104, 105, 106, 107, 108, 225, 226, 231
 Stake, R. E., 17, 31, 33, 147, 278
 Stanfield, J. H., II, 125, 126, 127
 Stanley, J. C., 8, 18, 19, 20, 21, 25, 62, 257, 258
 Steckler, A., 46
 Steiner, P. M., 63
 Stephen, N. A., 293
 Stevahn, L., 44, 45, 235, 237, 238, 239, 242
 Stiglitz, J. E., 111
 Stol, T., 244, 249
 Stufflebeam, D. L., 9, 18, 19, 37, 38, 39, 94, 266, 267
 Suchman, E. A., 25, 257
 Sweetman, D., 141
 Symonette, H., 127, 131, 151
 Takavarasha, K., 59, 60
 Tashakkori, A., 17
 Taut, S. M., 38
 Teddlie, C., 17
 Thomas, V. G., 127, 128, 229
 Thompson, D., 116
 Thompson-Robinson, M., 34, 145
 Tilley, N., 78
 Tillman, L. C., 34
 Tippet, L. H. C., 57, 64
 Tipton, E., 60
 Todman, J. B., 57
 Todrys, K. W., 139
 Torres, R. T., 43, 44, 74, 208, 227
 Triana-Tremain, B., 194
 Trickett, E. J., 34
 Trivette, C. M., 196
 Trochim, W. M. K., 299
 Tuck, E., 146, 148, 154
 Turchan, B. S., 57
 Turner, N., 82, 88
 Tversky, A., 187
 Tyler, R. W., 9, 18, 19, 24, 100, 225, 256
 Valentine, J. C., 59, 60
 Valls, R., 139
 van den Berg, R. D., 135, 189
 van 't Hof, S., 247
 Vidueira, P., 194
 Visse, M., 107
 Vo, A., 173, 176, 180
 Volkov, B. B., 44
 Wakiaga, L., 141
 Walker, R., 225
 Wallace, T., 212
 Wandersman, A., 193, 199, 204, 205, 208
 Ward Hood, D., 127, 131
 Watkins, J., 232
 Watkins, K., 232
 Waziyatawin, 154
 Weber, M., 16
 Weber-Pillwax, C., 153
 Wehipeihana, N., 244
 Weijer, C., 59
 Weiner, S., 170
 Weisinger, J. Y., 130
 Weiss, C. H., 7, 9, 15, 24, 25, 26, 27, 68, 73, 78, 85, 151, 173, 231, 261, 262
 Westine, C. D., 10, 78
 Whitmore, E., 42, 212, 215, 216, 217, 223, 231, 288
 Whitney, D., 232
 Wholey, J. S., 39, 118, 267, 268
 Widdershoven, G. A., 107
 Widianingsih, I., 139
 Willet, J. B., 58
 Williams, B., 232, 247
 Williams, D. D., 7
 Williams, J. E., 8
 Wilson, A. T., 46, 135, 136, 137, 140
 Wilson, S., 149
 Wilson, W. J., 170
 Witte, A. D., 21
 Woelders, S., 107
 Wolfe, S. M., 137
 Wong, V. C., 63
 Worley, J., 139
 Worthen, B. R., 8, 9
 Wothke, W., 57
 Wright, J., 260, 261
 Wright, S. R., 25
 Yang, K. W., 146, 148, 154
 Yellow Bird, M., 154
 Yeste, C., 139
 Young, M. D., 125
 Zimmer, R., 69
 Zimmerman, B., 232
 Zimmerman, M. A., 195, 196
 Zuberi, T., 125
 Zucman, G., 111
 Zukowski, A., 215

Subject Index

Note. *f* or *t* following a page number indicates a figure or a table.

- Accountability
 - developmental evaluation and, 248–249
 - empowerment evaluation and, 199, 204–205
 - practical policy evaluation and, 162–163
 - social accountability, 12, 14, 15–16
- Accuracy standards, 38
- Action model, 79–80, 277
- Action model–change model schema
 - application of, 80–82, 81*f*, 83*f*, 84*f*
 - clarification of program theory and, 80–82, 81*f*, 83*f*, 84*f*
 - theory-driven evaluation and, 79–80, 81*f*, 83*f*, 84*f*
 - transformative evaluation and, 142
- Active–reactive–interactive–adaptive evaluators, 41. *See also* Evaluator
- Adaptation, 246–247, 249–250
- Adaptive evaluation, 10, 187, 246–247, 248–249
- Addition programs, 95–96, 97
- Alignment patterns, 197–198, 198*f*
- Alkin, Marvin, 40–41
- Alliances, 251
- Altruistic point of view, 96–99
- Amistad* Supreme Court case, 129–132
- Anti-racist approaches, 142. *See also* Racial framing
- Appreciative inquiry (AI), 44, 232, 289–290
- Assumptions, 229–231. *See also* Bias
- Attrition, 63
- Axiological context, 153, 153*t*
- Backward reasoning, 80, 81*f*, 82, 83*f*, 84*f*, 277
- Baseline equivalence, 63
- Behavioral change, 73–74
- Behavioral theory, 88–89
- Bias
 - communicentric bias, 125–126
 - culturally responsive evaluation and, 125–128
 - deliberative democratic evaluation and, 114, 115, 116
 - Indigenous evaluation and, 151
 - learning-centered evaluation and, 232
 - systemic review and, 60
- Black, Indigenous, and People of Color (BIPOC) scholarship. *See also* Indigenous evaluation; Racial framing
- history of evaluation as a field or practice and, 3–4
- prescriptive theory development and, 7
- transformative evaluation and, 136–137, 138–139, 281–282
- Blame, 104–105
- Blue Marble evaluation, 188–189, 191. *See also* Utilization-focused evaluation (U-FE)
- Boruch, Robert, 20–21
- Bottom-up approach, 89, 297–298
- Brown v. Board of Education*, 128–129
- Campbell, Donald, 19–20, 257
- Capacity building
 - developmental evaluation and, 251
 - empowerment evaluation and, 199, 204, 207
 - evaluation theory–policy analysis and, 296–297
- Category systems, 8–9
- Causal mechanisms, 79, 257–259, 276–278
- Center for the Study of Evaluation (CSE) model, 9, 40–41
- Change and change models, 73–74, 79–80, 88–89, 135–142, 277. *See also* Societal changes
- Characteristic causal chain, 29
- Checklists, 168–169
- Chelimsky, Eleanor, 39–40
- Chen, Huey-Tsyh, 23–24
- Choice, 98–99
- Chouinard, Jill, 35–36
- Civil Rights Movement, 131
- Classical social theory, 16
- Classification systems, 8–9
- Clearinghouses, 61–62, 61*t*
- Cluster randomized trials, 57, 58–59. *See also* Randomized experiments
- Co-creation, 245, 246*f*
- Cognitive empathy, 112–113, 115, 279–280
- Collaboration
 - developmental evaluation and, 251
 - Indigenous evaluation and, 150–151, 153

- Collaboration (*cont.*)
 interactive evaluation practice and, 238
 learning-centered evaluation and, 231
 practical participatory collaborative evaluation and, 42–43
 theory-driven evaluation and, 88
 transformative evaluation and, 142
 Collaborative approaches to evaluation (CAE)
 overview, 212–213, 215–219, 218*f*, 221–223, 288–289
 in practice, 219–221, 220*t*
 principles, 214, 217–219, 218*f*, 220*t*
 See also Practical participatory evaluation (PPE)
 Colonialism, 145–147, 146*t*, 154. *See also* Indigenous evaluation
 Communication, 153–154
 Communicentric bias, 125–126. *See also* Bias
 Community context, 176, 199, 200, 204
 Community partners, 79–80, 150, 251
 Community-based organizations (CBOs), 89
 Competency, 187, 236, 238
 Competitive elaboration, 22–23
 Complexity theory, 186–187, 245, 246*f*, 247, 249
 Compliance, 184
 Comprehensive evaluation, 13*f*, 25–26, 260–261
 Conceptual framework, 26, 214
 Conflict management and resolution, 238
 Confounds, 62–63
 Construct validity, 24–25, 75
 Constructive methodology, 264–265
 Constructivism, 17–18
 Constructivist paradigm, 36–37
 Context. *See also* Context-sensitive evaluation (CSE); Cultural context
 context awareness, 176
 context-sensitive evaluation and, 175, 177*t*–180*t*
 developmental evaluation and, 249
 learning-centered evaluation and, 230
 overview, 94–95
 quasi-experimentation and, 258
 randomized experiments and, 59
 theory and, 80, 273
 utilization-focused evaluation (U-FE) and, 184
 Context, Input, Process, Product (CIPP) model, 8–9, 13*f*, 15, 37, 38–39, 94, 266–267
 Context Checkpoint, 96
 Context evaluation, 38, 266–267. *See also* Context, Input, Process, Product (CIPP) model
 Context-sensitive evaluation (CSE)
 activity sequence, 175
 evaluation theory tree and, 13*f*
 evidence-based policy use and, 261–262
 goal of, 174–175
 overview, 40–41, 173–174, 285
 path to observable actions, 177*t*–180*t*, 180–182
 primary applicability, 174–175
 research-based observable actions, 176, 177*t*–180*t*, 180
 Contextual analysis, 141
 Contextual support, 80
 Continuation decisions, 193
 Contractualism, 32
 Contribution analysis, 215
 Controlling for bias. *See* Bias
 Cook, Thomas, 21–22, 257–259
 Cooperative approaches, 239
 Core values. *See* Deep values; Values
 Cost analyses, 13*f*, 14, 30, 262–263
 Cost–benefit analyses, 262–263
 Cost-effectiveness analyses, 262–263
 Cousins, J. Bradley, 42–43
 Cram, Fiona, 35–36
 Credibility
 context-sensitive evaluation and, 175
 evaluation theory–policy analysis and, 293
 implications of a legislative context for evaluation practice and, 163
 practical policy evaluation and, 165–165, 284
 Criminal justice system, 139
 Crisis, 249–250
 Critical description, 29–30
 Critical friends, 199, 206
 Critical race theory, 138
 Critical realism, 17
 Cronbach, Lee J., 24–25, 259–260
 Cross-branch politics, 163–165. *See also* Policy
 CSE Activity Sequence, 175
 Cultural context. *See also* Context; Culturally responsive evaluation (CRE); Indigenous evaluation
 context-sensitive evaluation and, 177*t*–180*t*, 181–182
 equitable evaluation and, 265–266
 history of evaluation as a field or practice and, 3–4
 Culturally responsive equitable evaluation (CREE), 138–139
 Culturally responsive evaluation (CRE)
 collaborative approaches to evaluation (CAE) and, 216
 communicentric bias of evaluation knowledge production and, 125–126
 defining and situating, 128–131
 evaluation policy and, 292–297, 298
 evaluation theory tree and, 13*f*
 evaluation theory–policy analysis and, 292–297
 history of, 147
 overview, 33–34, 107, 126–128, 281
 purposes of, 131–132
 theory-driven evaluation and, 87
 transformative evaluation and, 138, 142
 white supremacy roots in evaluation and, 125–128
 See also Indigenous evaluation; Responsive evaluation
 Cultures of evidence, 199
 Cycles of reflection and action, 199–200
 Databases, 61–62, 61*t*
 Decision accretion, 27
 Decision making
 decision facilitation and management, 9
 deliberative democratic evaluation and, 116
 evaluation for improving social systems and, 24
 evidence-based policy use and, 262
 practical participatory evaluation and, 221–222
 transformative evaluation and, 34–35
 Decision-oriented theories, 37–38. *See also* Use branch of the evaluation theory tree
 Decolonization of evaluation, 147–148, 154
 Deductive testing of theory, 213
 Deep stories, 113–114, 116
 Deep values, 113–114, 116
 Deliberative democratic evaluation
 biases and, 114
 cultivation of cognitive empathy, 112–113, 115
 deep stories and, 113, 116
 deep values and, 113–114, 116
 evaluation theory tree and, 13*f*
 evaluators acting as moral fiduciaries and, 111–112, 115, 116
 implementation, 115–116
 learning-centered evaluation and, 231
 overview, 31–33, 110, 279–280
 racial and sexual framing and, 114
 responsive evaluation and, 31
 societal changes and, 110–111
 transparency and, 115, 116
 Deliberative justification, 33
 Democratic evaluation
 collaborative approaches to evaluation (CAE) and, 215
 equitable evaluation and, 265–266
 implications of a legislative context for evaluation practice and, 163–165

- learning-centered evaluation and, 231
 - See also* Deliberative democratic evaluation; Values-engaged democratic evaluation
- Democratic participation, 199
- Description theory, 6
- Descriptive assumptions, 78–79
- Development perspective, 162–163, 175, 246*f*
- Developmental evaluation (DE).
 - See also* Utilization-focused evaluation (U-FE)
 - context for, 249
 - evaluation theory tree and, 13*f*
 - history of evaluation as a field or practice and, 244
 - overview, 45, 188, 189*f*, 191, 243–244, 249–251, 291
 - purposes and principles of, 245–251, 246*f*
- Dialogical context, 153–154
- Discrimination, 137
- Dissemination, 89, 230–231, 296
- Distrust, 110–112, 115, 116
- Diversity, 125–132
- Documentation, 168, 173
- Drug education programs, 95–96, 97
- Ecological context, 80, 150
- Economic development, 139
- Educational connoisseurship, 13*f*, 29–30, 263–264
- Educational policy, 139
- Educational program evaluation
 - alternative pathways to evaluation influence and, 73
 - cost analyses and, 262–263
 - objectives-based evaluation and, 256–257
 - overview, 6–7, 19
 - transformative evaluation and, 141
 - values-engaged democratic evaluation and, 118, 122–124
 - See also* individual evaluation methods
- Egoistic point of view, 96–99
- Eight-Year Study, 18–19
- Eisner, Elliott, 29–30, 263–264
- Elaboration, 73–74
- Elimination decisions, 193
- Emancipatory justification, 33
- Emergent realist evaluation (ERE), 13*f*, 22–23, 275–276
- Empathy, 96–99, 112–113, 115
- Empiricism, 222–223
- Empowerment evaluation
 - collaborative approaches to evaluation (CAE) and, 215
 - concepts, 199–200
 - essential features, 205–207
 - evaluation theory tree and, 13*f*
 - Indigenous evaluation and, 36
 - overview, 41–42, 193–194, 207–208, 286–287
 - in practice, 194–203, 198*f*, 200*f*, 201*f*, 202*f*, 203*f*
 - principles, 198–199, 204, 206
 - theory and, 194–198, 198*f*, 203–205
 - three-step approach to, 200–203, 200*f*, 201*f*, 202*f*, 203*f*, 206–207
- Enlightenment use, 26
- Environmental policy, 139
- Environmental sustainability, 189–190, 191
- Epistemology
 - evaluation influence and, 75–76
 - evaluation theory tree and, 12, 13*f*, 14
 - Indigenous evaluation and, 149–150, 151
 - overview, 17–18
- Equitable evaluation
 - evaluation theory tree and, 13*f*
 - interactive evaluation practice and, 240–241
 - overview, 37, 265–266
 - transformative evaluation and, 138, 142
- Equity, 4–5, 189–190, 191, 228–229
- Ethics
 - collaborative approaches to evaluation (CAE) and, 216–217
 - context-sensitive evaluation and, 175
 - ethical fallacies, 32
 - evaluation theory–policy analysis and, 297
 - evaluators acting as moral fiduciaries and, 111–112
 - human work and, 104
 - Indigenous evaluation and, 153, 153*t*
 - learning-centered evaluation and, 230
 - overview, 99
 - randomized experiments and, 58–59
 - responsive evaluation and, 102
 - transformative evaluation and, 34, 136, 137–138, 141–142
- Ethos, 104
- Evaluability assessment, 39, 267–268
- Evaluand, 14, 27–28, 102–103
- Evaluation design, 295
- Evaluation for improving social system, 13*f*, 24–25, 259–260
- Evaluation influence
 - alternative pathways to, 72–75
 - epistemological roots, 75–76
 - evaluation as assisted sensemaking, 67–69
 - overview, 23, 38, 67, 76
 - purposes of evaluation and, 71–72
 - social betterment and, 67, 69–71
 - See also* Outcomes
- Evaluation issue analysis, 7, 8
- Evaluation planning, 70, 71–72, 175
- Evaluation policy. *See* Policy
- Evaluation teaching and interpretation, 7, 8, 230–231, 263–264
- Evaluation theory
 - defining and situating CRE in, 128–131
 - evaluation policy and, 292–298
 - as a guide to practice, 9–10
 - legislative context and, 161–165
 - overview, 6–8, 76, 93–99, 255
 - purposes of evaluation and, 71–72
 - values-engaged democratic evaluation and, 120
 - See also* Evaluation theory tree
- Evaluation theory tree
 - branches of, 13*f*, 18–45
 - overview, 9, 12–15, 13*f*, 45–46, 255
 - practical participatory evaluation and, 213
 - roots of, 12–14, 13*f*, 15–18
 - See also* Evaluation theory; Methods branch of the evaluation theory tree; Use branch of the evaluation theory tree; Valuing branch of the evaluation theory tree
- Evaluation useability, 177*t*–180*t*, 295–296
- Evaluation-orientation research methodology, 7, 8
- Evaluative thinking
 - collaborative approaches to evaluation (CAE) and, 217–219, 218*f*, 220*t*, 221
 - learning-centered evaluation and, 230
 - utilization-focused evaluation (U-FE) and, 187
- Evaluator
 - CIPP model and, 267
 - collaborative approaches to evaluation (CAE) and, 217, 220*t*, 221
 - context-sensitive evaluation and, 175, 177*t*–180*t*, 180–181
 - culturally responsive evaluation and, 131
 - educational connoisseurship and, 30, 264
 - evaluators acting as moral fiduciaries and, 111–112, 115, 116
 - fourth-generation evaluation and, 36
 - goal-free evaluation and, 28
 - Indigenous evaluation and, 152
 - interactive evaluation and, 44–45, 236–237, 239–240
 - learning-centered evaluation and, 43–44, 228, 230
 - practical participatory collaborative evaluation and, 42–43
 - practical participatory evaluation and, 221–222

- Evaluator (*cont.*)
 practical policy evaluation and, 165, 169–171
 transformative evaluation and, 34–35
 utilization-focused evaluation (UFE) and, 41
- Evidence Act, 292–293
- Evidence-based interventions (EBIs), 89
- Evidence-based policy use, 13/
 26–27, 261–262
- Evidence-based strategies, 199, 215, 217
- Exhibited generalization approach, 87–88, 277–278
- Experimental methods
 evaluation theory tree and, 13/
 goal-free evaluation and, 28–29
 overview, 19–20, 25–26, 257
 results-oriented evaluation and, 268
 social inquiry and, 16
- External validity
 evaluation influence and, 75
 overview, 19
 tailored evaluation and, 26
 theory-driven evaluation and, 24, 87–88
- Failure analysis, 63–64
- Feasibility standards, 38
- Feedback, 245, 246/*f*, 248, 267, 268
- Feminist approaches, 142
- Fetterman, David, 41–42
- Fidelity challenge, 245
- Fiduciary, moral. *See* Moral fiduciaries
- Flexibility, 170–171
- Forces Framework, 166–168. *See also* Practical policy evaluation
- Formalization, 101–102
- Formative evaluation
 context-sensitive evaluation and, 173, 174
 developmental evaluation and, 245–246
 overview, 10
 theory-driven evaluation and, 86–87
 values-engaged democratic evaluation and, 122
- Forward reasoning, 81/*f*, 82, 83/*f*, 84/*f*, 277
- Forward–backward reasoning, 81/*f*, 82, 84/*f*, 277
- Fourth-generation evaluation, 13/*f*, 36–37, 264–265
- Free-range evaluation, 236
- Gender, 139, 140–141, 279–280. *See also* LGBTQI community
- Generalization of results, 59–60, 87–88, 258
- Globalization, 186, 249
- Goal-free evaluation, 13/*f*, 28–29, 94
- Goals
 accountability and, 15
 collaborative approaches to evaluation (CAE) and, 215–216
 context-sensitive evaluation and, 174–175
 evaluation theory–policy analysis and, 294
 interactive evaluation practice and, 235–236
 learning-centered evaluation and, 230
 purposes of evaluation and, 71–72
 theory and, 10
 theory-driven evaluation and, 79, 85/*t*
 values-engaged democratic evaluation and, 120–121
See also Outcomes; Social betterment goal
- Go/no-go decisions, 24, 193
- Government
 distrust and, 110–111
 implications of a legislative context for evaluation practice, 163–165
 implications of a legislative context for evaluation practice and, 169
 practical policy evaluation and, 163–165, 283–285
- Government Accountability Office (GAO), 16
- Greene, Jennifer C., 33
- Grounded theory, 205
- Group randomized experiments, 57, 62. *See also* Randomized experiments
- Guba, Egon, 36–37, 264–265
- Guide to practice, 9–10, 118–119, 187, 189–190
- Guiding star. *See* Social betterment goal
- Henry, Gary, 22–23
- Hermeneutic circle, 36
- Heterogeneity, 142
- Heuristics, 187
- Hilliard, Asa, 37, 265–266
- History of evaluation as a field of practice
 developmental evaluation and, 244
 Indigenous evaluation, 147–148
 overview, 3–5, 93–94
 practical policy evaluation and, 161–165
 utilization-focused evaluation (U-FE) and, 184–187
 white supremacy roots in evaluation and, 125–128
- Holistic effectuality evaluation, 87
- Hopson, Rodney, 33–34
- House, Ernest, 31–33
- Human factors, 88
- Humanism, 103–104
- Humanitarian aid, 245
- Humility, 237
- Humor, 237
- Identity, 80
- Implementation
 collaborative approaches to evaluation (CAE) and, 219–221, 220/*t*
 cost analyses and, 262–263
 deliberative democratic evaluation and, 115–116
 empowerment evaluation and, 194, 200–203, 200/*f*, 201/*f*, 202/*f*, 203/*f*, 205–207
 interactive evaluation practice and, 239–240
 learning-centered evaluation and, 229–230
 practical participatory evaluation and, 213–214
 theory-driven evaluation and, 82, 85–86, 85/*t*
 utilization-focused evaluation (U-FE) and, 187
See also Practice
- Improvement, 193, 199, 246–247
- Inclusion, 116, 136, 199
- Inclusive/transformative evaluation, 34–35. *See also* Transformative evaluation
- Indian Health Service (IHS), 292–298
- Indigenous evaluation
 evaluation theory tree and, 13/*f*
 evaluation theory–policy analysis and, 292–297
 history of, 147–148
 locating within CRE practice, 148–154, 153/*t*
 overview, 35–36, 145–147, 146/*t*, 154–155, 282–283
 transformative evaluation and, 142
See also Culturally responsive evaluation (CRE)
- Indigenous sovereignty, 146–147, 146/*t*. *See also* Indigenous evaluation
- Individual context, 176, 181–182
- Inductive theory building, 213
- Inequity
 evaluators acting as moral fiduciaries and, 111–112
 interactive evaluation practice and, 240–241
 societal changes and, 110–111
 transformative evaluation and, 135, 281–282
- Influence, 27, 176
- Innovation
 developmental evaluation and, 245, 246/*f*, 247, 249, 251
 social innovation, 243–244
 utilization-focused evaluation (U-FE) and, 187

- Input evaluation, 38, 266–267. *See also* Context, Input, Process, Product (CIPP) model
- Institutional context, 111, 152–153
- Intensive evaluation, 39, 267, 268
- Interactive evaluation, 13*f*, 44–45, 238
- Interactive evaluation practice (IEP)
goals of, 235–236
implementation, 239–240
overview, 44–45, 234–237, 241–242, 290–291
positionality and, 240–241
principles, 237–238
settings for, 236
- Internal validity
overview, 19
quasi-experimentation and, 258
tailored evaluation and, 26
theory-driven evaluation and, 24, 87
- Interorganization collaboration, 88
- Interpersonal skills, 235, 236–238
- Interpretation, 187
- Intersectionality, 142
- Interventions, 79, 80, 88–89, 276–278. *See also* Treatments
- Judgment
developmental evaluation and, 244
educational connoisseurship and, 264
goal-free evaluation and, 28–29
learning-centered evaluation and, 231
social betterment and, 67–68
utilization-focused evaluation (U-FE) and, 184
- Julnes, George, 22–23
- Justice in evaluation, 7, 32, 110. *See also* Deliberative democratic evaluation
- Key Evaluation Checklist (KEC)
model, 94–95, 278
- King, Jean, 44–45
- Knowledge perspective, 162–163
- Language, 153–154
- Learning, 248–249
- Learning-centered evaluation
assumptions and, 229–231
empowerment evaluation and, 197
evaluation theory tree and, 13*f*
overview, 43–44, 225–233, 289–290
- Legislative considerations
collaborative approaches to evaluation (CAE) and, 216
implications of for evaluation practice, 163–165, 169
practical policy evaluation and, 163–165, 283–285
randomized experiments and, 58–59
- Levin, Henry, 30, 262–263
- LGBTQI community, 140–141. *See also* Sexual framing
- Liberation, 131–132
- Lincoln, Yvonna, 36–37, 264–265
- Lived experience, 130, 155
- Logic model, 175
- Longitudinal study, 58–59
- Macrolevel contextual support, 80
- Management of evaluations, 60, 294
- Management or development
perspective, 32, 162–163, 175
- Marginalized communities, 240–241, 279–280, 281–282. *See also* Black, Indigenous, and People of Color (BIPOC) scholarship; Indigenous evaluation; Racial framing; Sexual framing
- Mark, Melvin, 22–23
- Measures, 175
- Mediators, 59
- Mental health, 139
- Mertens, Donna, 34–35
- Meta-analysis, 60–63, 61*t*, 73, 188
- Meta-evaluation, 101–102, 258–259
- Metatheory of evaluation, 93–99
- Methods and methodology
deliberative democratic evaluation and, 32
developmental evaluation and, 247–248
evaluation theory–policy analysis and, 295
Indigenous evaluation and, 150–151
practical policy evaluation and, 165–166
social betterment as a goal and, 70–71
social inquiry and, 16
transformative evaluation and, 137–138
- Methods branch of the evaluation theory tree, 13*f*, 14–15, 18–27, 256–262. *See also* Comprehensive evaluation; Context, Input, Process, Product (CIPP) model; Educational connoisseurship; Emergent realist evaluation (ERE); Evaluation for improving social system; Evaluation theory tree; Evidence-based policy use; Experimental methods; Objectives-based evaluation; Quasi-experimentation; Randomized experiments; Results-oriented evaluation; Theory-driven evaluation (TDE)
- Microlevel contextual support, 80
- Missing data, 63
- Mixed-methods designs, 85–86, 85*t*, 138
- Moderators, 59
- Modus operandi (MO) method, 28–29
- Moral fiduciaries, 111–112, 115, 116, 279–280
- Motivations, 96–99
- Multidisciplinary model, 227–228
- Multilateral self-awareness, 151–152
- Multiwave formative evaluation, 86
- Needs assessment, 28
- Networks, 251
- Objectives-based evaluation, 13*f*, 18–19, 256–257
- Objectivist thinking, 14, 27–28
- Objectivity, 18, 103
- Observable actions, 10, 176, 177*t*–180*t*, 180–182. *See also* Theory prototypes
- Observations, 24, 187, 259–260. *See also* UTOS (units, treatments, observations, and settings) acronym
- Ontology, 18, 149, 151
- Oppression, systems of, 125–132, 137, 142
- Organizational context
associated organizations, 79–80
context-sensitive evaluation and, 175, 176, 181–182
developmental evaluation and, 251
empowerment evaluation and, 199
learning-centered evaluation and, 230, 232
- Outcomes
alternative pathways to evaluation influence, 72–75
context-sensitive evaluation and, 173
cost analyses and, 263
evaluation for improving social systems and, 24
goal-free evaluation and, 28–29
learning-centered evaluation and, 230–231
monitoring, 39, 267, 268
purposes of evaluation and, 71–72
randomized experiments and, 62
responsive evaluation and, 105–107
theory-driven evaluation and, 79, 85*t*, 87
See also Evaluation influence; Goals Outgroup homogeneity bias, 113
- Paradigm, 17
- Participation
context-sensitive evaluation and, 176, 177*t*–180*t*, 181
culturally responsive evaluation and, 129–130
equitable evaluation and, 265–266
evaluation theory–policy analysis and, 294–295
See also Participatory forms of evaluation

- Participatory action research, 215
- Participatory forms of evaluation
- collaborative approaches to evaluation (CAE) and, 217–219, 218*f*, 220*t*
 - Indigenous evaluation and, 36
 - interactive evaluation practice and, 235–236, 237–238
 - learning-centered evaluation and, 231
 - transformative evaluation and, 142
- See also* Participation
- Patterns of alignment, 197–198, 198*f*
- Patton, Michael, 41, 45
- Performance monitoring, 39, 267, 268
- Personal context, 151–152
- Philosophical rationale, 7, 216
- Pivot, 249–250
- Planning evaluations, 70, 71–72, 175
- Pluralism/elitism, 32
- Policy
- considering past history and, 166–168
 - developing an evaluation policy, 297–298
 - evaluation for improving social systems and, 24
 - evaluation theory–policy analysis and, 292–297
 - evidence-based policy use and, 26–27, 261–262
 - implications of a legislative context for evaluation practice and, 163–165
 - overview, 298
 - social betterment and, 70
 - sustainability and equity and, 189–190
 - transformative evaluation and, 139
- See also* Practical policy evaluation
- Political context
- collaborative approaches to evaluation (CAE) and, 216
 - context-sensitive evaluation and, 175
 - implications of a legislative context for evaluation practice and, 163–165, 169
 - Indigenous evaluation and, 151
 - learning-centered evaluation and, 230, 231
 - results-oriented evaluation and, 267
- Populations, 24, 80, 111–112
- Positionality, 240–241
- Positivism, 17
- Postpositivism, 17
- Power, 142, 231
- Practical participatory collaborative evaluation, 13*f*, 42–43
- Practical participatory evaluation (PPE), 211–214, 221–222, 288–289. *See also* Collaborative approaches to evaluation (CAE)
- Practical policy evaluation
- assumptions regarding the use of findings and, 162–163
 - checklists and, 168–169
 - considering past history and, 166–168
 - developing credibility defenses, 165–166
 - evaluation theory tree and, 13*f*
 - history of evaluation practice and, 161–165
 - implications of a legislative context for evaluation practice and, 163–165, 169
 - overview, 39–40, 161, 169–171, 283–285
 - questions and, 168–169
 - user perspectives, 169
- See also* Policy
- Practice, 9–10, 229–230, 245–247. *See also* Implementation
- Pragmatism, 17, 18, 33, 216, 288
- Preformative evaluation, 245
- Preordinate evaluation, 28
- Prescriptions, 10, 78–79
- Prescriptive theories
- interactive evaluation practice and, 235
 - overview, 6–7, 173–174, 273
 - practical participatory evaluation and, 214
- See also* Context-sensitive evaluation (CSE); Theory prototypes
- Presentational effectiveness, 165, 166
- Preskill, Hallie, 43–44
- Principled-focused evaluation
- Blue Marble evaluation, 188–189
 - developmental evaluation and, 245–249, 246*f*
 - overview, 23, 188, 189*f*, 191
- See also* Utilization-focused evaluation (U-FE)
- Privilege, 240–241
- Proactive evaluation, 86–87
- Process evaluation, 38, 173, 185–186, 266–267. *See also* Context, Input, Process, Product (CIPP) model; Theory-driven evaluation (TDE)
- Process use, 195, 196–197, 203, 230
- Product evaluation, 38, 267. *See also* Context, Input, Process, Product (CIPP) model
- Program changes, 267–268
- Program evaluation. *See* Educational program evaluation; Social program evaluation; *individual evaluation methods*
- Program Evaluation and Methodology Division (PEMD)
- changes to evaluation practice in (1980–1994), 165–166
 - considering past history and, 166–168
 - overview, 163, 169–170
 - questions and the use of checklists, 168–169
- See also* Practical policy evaluation
- Program implementers, 79
- Program stress test, 86–87
- Program theory. *See also* Theory-driven evaluation (TDE)
- clarifying, 80–82, 81*f*, 83*f*, 84*f*
 - theory-driven evaluation and, 78–82, 81*f*, 83*f*, 84*f*
- Purposes of evaluation
- culturally responsive evaluation and, 131–132
 - developmental evaluation and, 245–251, 246*f*
 - evaluation theory–policy analysis and, 294
 - learning-centered evaluation and, 229
 - overview, 71–72, 76
- See also* Evaluation planning; Goals; Outcomes; Social betterment goal
- Quality control, 60–63, 61*t*
- Quasi-experimental designs (QEDs), 62, 63
- Quasi-experimentation
- evaluation theory tree and, 13*f*
 - goal-free evaluation and, 28–29
 - overview, 19, 20, 21–22, 257–259
 - results-oriented evaluation and, 268
- Racial framing
- culturally responsive evaluation and, 125–131
 - deliberative democratic evaluation and, 114, 115, 279–280
 - developmental evaluation and, 250–251
 - history of evaluation as a field or practice and, 3–4
 - learning-centered evaluation and, 228–229
 - transformative evaluation and, 135–142
- See also* Black, Indigenous, and People of Color (BIPOC) scholarship
- Randomization, 58–59
- Randomized control trials (RCTs), 55–56, 56*f*, 70–71, 73
- Randomized experiments
- ethical and legal considerations, 58–59
 - evaluation theory tree and, 13*f*
 - experimental methods and, 257
 - failure analysis and, 63–64
 - management of, 60
 - overview, 20–21, 55–56, 56*f*, 64
 - replication and generalization and, 59–60
 - statistical theory and operationalization, 56–58
 - systemic review and, 60–63, 61*t*

- Randomized field experiments, 20–22, 55–56, 56*f*, 274–275
- Rapid-feedback evaluation, 39, 267, 268
- Rational approach, 98–99
- Realism, 17, 27–28, 75
- Realistic evaluations. *See* Theory-driven evaluation (TDE)
- Reciprocity, 153, 153*t*, 216
- Recognition, 216
- Recruitment, 80
- Reflective processes, 142, 199, 200, 207, 216, 229
- Refugees, 139
- Relational context
- collaborative approaches to evaluation (CAE) and, 216, 217–219, 218*f*, 220*t*, 221
 - Indigenous evaluation and, 152
 - interactive evaluation practice and, 237–238
- Relativism, 17–18, 27–28, 32
- Replication of results, 59–60
- Reporting results or progress, 105–107, 230–231, 295–297
- Research, 7–8, 10
- Research-based observable actions, 176, 177*t*–180*t*, 180. *See also* Observable actions
- Resource availability
- collaborative approaches to evaluation (CAE) and, 217–219, 218*f*, 220*t*
 - cost analyses and, 262–263
 - results-oriented evaluation and, 267
- Respect, 153, 153*t*, 175, 237
- Response, 101–102
- Responsibility
- collaborative approaches to evaluation (CAE) and, 216
 - equitable evaluation and, 265–266
 - Indigenous evaluation and, 153, 153*t*
 - interactive evaluation practice and, 238
- Responsive evaluation
- blame and, 104–105
 - branches of, 107–108
 - evaluation theory tree and, 13*f*
 - human work and, 103–104
 - judging values and, 102
 - overview, 31, 100, 278–279
 - responsivity, 102–103
 - stories and, 105–107
 - valuings, 100–101
 - See also* Culturally responsive evaluation (CRE)
- Responsiveness, 131
- Results-oriented evaluation, 13*f*, 39, 184, 267–268
- Rights, 216
- Rigorous studies
- developmental evaluation and, 245, 246*f*
 - evaluation theory–policy analysis and, 295
 - evidence-based policy use and, 261–262
 - overview, 14
 - randomized experiments and, 62
- Role of psychologists in empowerment, 196, 206
- Roots of the evaluation theory tree, 12–14, 13*f*, 15–18. *See also* Epistemology; Evaluation theory tree; Social accountability; Social inquiry
- Rossi, Peter, 25–26, 260–261
- Scaling principles, 245
- Screening, 80
- Scriven, Michael, 28–29
- Self-determinism, 35, 41–42, 193, 195, 196
- Sensemaking, 67–68, 76, 230–231
- Service, 80
- Settings, 24, 259–260. *See also* UTOS (units, treatments, observations, and settings) acronym
- Sexual framing
- deliberative democratic evaluation and, 114, 115, 279–280
 - learning-centered evaluation and, 228–229
 - transformative evaluation and, 140–141
 - See also* LGBTQI community
- Social accountability, 12, 14, 15–16
- Social betterment goal
- alternative pathways to evaluation influence and, 73
 - emergent realist evaluation (ERE) and, 274–275
 - overview, 67, 69–71, 76
 - See also* Evaluation influence
- Social context, 114, 175, 265–266, 274–275. *See also* Bias
- Social innovation, 243–244
- Social inquiry, 12, 13*f*, 14, 16–17
- Social justice
- cultivation of cognitive empathy, 112–113
 - deliberative democratic evaluation and, 110–111
 - empowerment evaluation and, 199
 - evaluators acting as moral fiduciaries and, 111–112
 - learning-centered evaluation and, 228–229, 231
 - theories and, 7
 - transformative evaluation and, 281–282
- Social program evaluation, 118, 122–124. *See also individual evaluation methods*
- Social science inquiry, 7, 23
- Social theory, 16
- Societal changes
- deliberative democratic evaluation and, 110–111
 - developmental evaluation and, 249
 - evaluators acting as moral fiduciaries and, 111–112
 - transformative evaluation and, 135–142, 281–282
 - See also* Systems change
- Stake, Robert, 31
- Stakeholder theory, 88–89
- Stakeholders
- biases and, 114
 - clarification of program theory and, 80–82, 81*f*, 83*f*, 84*f*
 - collaborative approaches to evaluation (CAE) and, 217–219, 218*f*, 220*t*, 221
 - Context, Input, Process, Product (CIPP) model and, 38–39
 - context-sensitive evaluation and, 174, 175, 176, 177*t*–180*t*, 180–182
 - deliberative democratic evaluation and, 32, 116
 - emergent realist evaluation (ERE) and, 22
 - empowerment evaluation and, 197
 - equitable evaluation and, 266
 - evaluation as assisted sensemaking, 68
 - evidence-based policy use and, 262
 - fourth-generation evaluation and, 36, 264–265
 - goal-free evaluation and, 28
 - Indigenous evaluation and, 35–36, 147, 150–151
 - interactive evaluation practice and, 235, 236
 - learning-centered evaluation and, 43–44, 228, 229–230, 231
 - objectives-based evaluation and, 256
 - practical participatory evaluation and, 42–43, 221–222
 - prescriptive theory development and, 7
 - responsive evaluation and, 31, 101
 - results-oriented evaluation and, 267
 - social betterment and, 67–68
 - theory-driven evaluation and, 78–82, 81*f*, 83*f*, 84*f*
 - transformative evaluation and, 35, 136–137, 138, 142
 - utilization-focused evaluation (U-FE) and, 184, 185
 - values-engaged democratic evaluation and, 120, 121–123
- Standardization
- Context, Input, Process, Product (CIPP) model and, 38
 - context-sensitive evaluation and, 175
 - responsive evaluation and, 108
 - values-engaged democratic evaluation and, 119–120, 119*t*
- *UTOS (star-UTOS), 24, 259–260. *See also* UTOS (units, treatments, observations, and settings) acronym

- Statistical software, 57, 59–60
 Statistical theory, 56–58, 274–275
 Status quo, 228–229, 231
 Step wedge designs, 59
 Stereotyping, 114, 125–128
 Stories, 105–107, 113, 116
 Strategic evaluation, 228
 Stufflebeam, Daniel, 38–39, 266–267
 Subjective meaningfulness, 14
 Subjectivity, 18, 27–28, 31, 103
 Substantive/methodological quality, 165–166
 Summary formative evaluation, 10, 173, 174
 Summative evaluation, 10, 122, 173, 193
 Sustainability, 189–190, 191, 228
 Sustainable Development Goals (SDGs), 188
 Sympathetic motivation, 96–99
 Systematic research methods, 67, 69–71. *See also* Evaluation influence
 Systematic reviews, 60–63, 61*t*
 Systematic social inquiry. *See* Social inquiry
 Systems change. *See also* Societal changes
 developmental evaluation and, 245, 247, 249, 250–251
 learning-centered evaluation and, 228–229, 230
 Systems thinking, 186–187, 245, 246*f*

 Tailored evaluation, 25–26. *See also* Comprehensive evaluation
 Target group, 80, 85*t*. *See also* Populations
 Teaching, evaluation. *See* Evaluation teaching and interpretation
 Theories of use and action, 195
 Theory. *See also* Evaluation theory; Evaluation theory tree; Theory prototypes; Theory–practice relationship; *individual theory prototypes*
 assumptions regarding the use of findings and, 162–163
 category systems and, 8–9
 collaborative approaches to evaluation (CAE) and, 223
 deliberative democratic evaluation and, 116–117
 developing an evaluation policy and, 297–298
 developmental evaluation and, 245–247
 empowerment evaluation and, 194–198, 198*f*, 203–207
 evaluation theory–policy analysis and, 292–297
 as a guide to practice, 9–10
 learning-centered evaluation and, 226–227
 overview, 6–8, 10, 17, 93–99, 255, 273–274, 292
 practical participatory evaluation and, 213–214
 theory prototypes, 273–274
 Theory of action, 197–198, 198*f*, 204
 Theory of change, 78–82, 81*f*, 83*f*, 84*f*
 Theory of use, 197–198, 198*f*, 204
 Theory prototypes. *See* Theory; *individual theory prototypes*
 Theory-driven evaluation (TDE)
 application of, 88–89
 conventional approaches to, 82, 85–86, 85*t*
 evaluation theory tree and, 13*f*
 innovative approaches to, 86–88
 overview, 23–24, 78, 89, 276–278
 stakeholders and, 78–82, 81*f*, 83*f*, 84*f*
 See also Program theory
 Theory-driven outcome evaluation, 82
 Theory-driven process evaluation, 85–86, 85*t*
 Theory–practice relationship, 273–274, 292–297. *See also* Theory
 Three-step model in empowerment evaluation, 194, 200–203, 200*f*, 201*f*, 202*f*, 203*f*, 206–207, 287. *See also* Empowerment evaluation
 Timely feedback, 245, 246*f*
 Top-down approach, 89
 Transdiscipline, 29
 Transformational learning, 43–44
 Transformative evaluation
 applications of, 139
 assumptions of, 135–138
 collaborative approaches to evaluation (CAE) and, 215
 developmental evaluation and, 250–251
 evaluation theory tree and, 13*f*
 examples of, 140–142
 identifying a transformative theoretical approach, 141–142
 Indigenous evaluation and, 148
 overview, 34–35, 135, 140–142, 281–282
 utilization-focused evaluation (U-FE) and, 189–190
 Transformative participatory evaluation (TPE), 215, 216
 Transparency, 115, 116, 142
 Transscientific theory of evaluation, 93–99
 Treatments, 24, 79, 85*t*, 259–260. *See also* Interventions; UTOS (units, treatments, observations, and settings) acronym
 Trust
 context-sensitive evaluation and, 175
 evaluators acting as moral fiduciaries and, 111–112
 societal changes and, 110–111
 transformative evaluation and, 142
 transparency and, 115, 116
 Truth relativism, 18
 Tyler, Ralph, 18–19, 256–257
 Type of evaluation, 10. *See also* Adaptive evaluation; Formative evaluation; Summary formative evaluation; Summative evaluation

 Units, 24, 56–57, 259–260. *See also* UTOS (units, treatments, observations, and settings) acronym
 Usability of findings, 177*t*–180*t*, 295–296
 Use branch of the evaluation theory tree, 13*f*, 14–15, 37–45, 266–268. *See also* Context-sensitive evaluation (CSE); Developmental evaluation (DE); Empowerment evaluation; Evaluation theory tree; Interactive evaluation; Learning-centered evaluation; Practical participatory collaborative evaluation; Practical policy evaluation; Utilization-focused evaluation (U-FE)
 User context
 context-sensitive evaluation and, 176, 177*t*–180*t*, 181
 practical policy evaluation and, 169
 utilization-focused evaluation (U-FE) and, 184, 185
 Utility standards, 38, 103
 Utilization-focused evaluation (U-FE)
 collaborative approaches to evaluation (CAE) and, 215
 developmental evaluation and, 45, 245, 246*f*, 251
 evaluation theory tree and, 13*f*
 evolution of, 184–187
 facilitation principles and skills, 187
 interactive evaluation practice and, 235
 new evaluation approaches and, 187–190, 189*f*
 overview, 41, 183–184, 189*f*, 190–191, 285–286
 status and influence of, 184
 sustainability and equity and, 189–190, 191
 UTOS, 24, 259–260
 utos (units, treatments, observations, and settings) acronym, 24, 259–260

 Validity, 19, 24–25, 75. *See also* External validity; Internal validity

- Values
 - deep values, 113–114, 116
 - deliberative democratic evaluation
 - and, 113–114, 116
 - democratic values, 123–124
 - transformative evaluation and, 135–136
 - values inquiry, 69
- Values-engaged democratic evaluation
 - evaluation theory tree and, 13*f*
 - goals of, 120–121
 - overview, 33, 118–124, 119*t*, 280–281
 - in practice, 120–123
- Valuing branch of the evaluation theory tree, 13*f*, 14–15, 27–37, 262–266. *See also*
 - Cost analyses; Culturally responsive evaluation (CRE); Deliberative democratic evaluation; Equitable evaluation; Evaluation theory tree; Fourth-generation evaluation; Goal-free evaluation; Indigenous evaluation; Responsive evaluation; Transformative evaluation; Values-engaged democratic evaluation
- Valuings, 100–102
- Weiss, Carol, 26–27, 261–262
- Western evaluation paradigms, 150, 152. *See also* Culturally responsive evaluation (CRE)
- What Works Clearinghouse (WWC), 61*t*, 62–63
- Whole-problem complexity, 170
- Wholey, Joseph, 39, 267–268

About the Editors

Marvin C. Alkin, EdD, is Professor Emeritus in the Social Research Methodology Division of the School of Education and Information Studies at the University of California, Los Angeles (UCLA). He has written extensively on evaluation practice, evaluation utilization, and comparative evaluation theory. Dr. Alkin is a founder and former Director of the UCLA Center for the Study of Evaluation. He has been a consultant to six national governments and has conducted more than 85 evaluations of a variety of educational, governmental, and foundation programs. He is a recipient of the Lazarsfeld Evaluation Theory Award and the Research on Evaluation Award from the American Evaluation Association (AEA).

Christina A. Christie, PhD, is Wasserman Dean and Professor in the School of Education and Information Studies at UCLA. Her work has three main foci: applied evaluation research studies, research on evaluation practice, and theoretical analysis. Dr. Christie is a recipient of the Marcia Guttentag Early Career Award and the Research on Evaluation Award from the AEA, as well as the Distinguished Scholar Award from the Research on Evaluation Special Interest Group of the American Educational Research Association. She has served on the board of the AEA and as an associate editor and section editor of the *American Journal of Evaluation*.

Contributors

Hind Al Hudib, PhD, is an independent researcher who completed a postdoctoral fellowship at the Centre for Research on Educational and Community Services, University of Ottawa.

Marvin C. Alkin, EdD (see “About the Editors”).

Robert F. Boruch, PhD, is a University Trustee Chair Professor at the University of Pennsylvania in the Graduate School of Education and the Statistics Department of the Wharton School and is a member of the graduate group for the Criminology Department. He is a founding member of the Evaluation Research Society, a progenitor of the American Evaluation Association, and the International Campbell Collaboration.

Eleanor Chelimsky, PhD (deceased), was a consultant in evaluation methodology, practice, and policy. She was Assistant Comptroller General of the U.S. Government Accountability Office for Evaluation and Methodology and, for 14 years, ran a division of about 100 social scientists who performed evaluations for the Congress in a very wide variety of subject areas. She served as president of the Evaluation Research Society (1981) and the American Evaluation Association (1995).

Huey T. Chen, PhD, is Professor in the Department of Public Health and Director of the Center for Evaluation and Applied Research at Mercer University. His previous positions included serving as a branch chief and senior evaluation scientist at the Centers for Disease Control and Prevention.

Jill Anne Chouinard, PhD, is Associate Professor in the School of Public Administration, University of Victoria, where she teaches, practices, and writes about the practice of evaluation. Much of her evaluation work has been conducted in culturally and socially diverse community settings, where she has extensive experience leading evaluations in the areas of education, social services, public health, community development, and organizational learning and change.

Christina A. Christie, PhD (see “About the Editors”).

J. Bradley Cousins, PhD, is Professor Emeritus and former Director of the Centre for Research on Educational and Community Services, University of Ottawa. He is also a Distinguished Visiting Professor at Claremont Graduate University.

Fiona Cram, PhD, is Director at Katoa Ltd. in Aotearoa New Zealand. She has over 25 years of Kaupapa Māori (by, with, and for Māori) research and evaluation experience with Māori and Iwi organizations and communities, as well as with government agencies, district health boards, and philanthropic organizations.

Elizabeth FastHorse, MA, is a doctoral candidate in the School of Education–Social Sciences and Comparative Education at the University of California, Los Angeles. Her research focuses on storytelling with the Rincon Youth Storytellers on the Rincon Indian reservation to preserve language and culture and use a tribal critical race theory method.

David M. Fetterman, PhD, is president and chief executive officer of Fetterman & Associates, an international evaluation consulting firm, and is concurrently a faculty member at Claremont Graduate University, Pacifica Graduate Institute, and the University of Arkansas. He is a past president of the American Evaluation Association and the American Anthropological Association’s Council on Anthropology and Education.

Emi Fujita-Conrads, MA, is an evaluator at Education Northwest, a nonprofit research organization in Portland, Oregon. Her research interests focus on methods for conducting social justice–focused and culturally responsive evaluations and evaluation policies.

Jennifer C. Greene, PhD, is Professor Emerita of Educational Psychology at the University of Illinois at Urbana–Champaign. She was the 2011 president of the American Evaluation Association.

Gary T. Henry, PhD, is Dean of the University of Delaware’s College of Education and Human Development and Professor in the School of Education and the Joseph R. Biden, Jr. School of Public Policy and Administration. A respected researcher in the field of education, he specializes in education policy, educational evaluation, educator labor markets, and quantitative research methods.

Jennifer E. Ho, PhD, is Deputy Director of Impact Assessment and Learning at the James Irvine Foundation. Her past and present work with international nonprofits, academia, and philanthropy focuses on the development and use of evaluation and strategic learning approaches in the service of historically marginalized communities.

Rodney Hopson, PhD, is Professor of Evaluation in the Department of Educational Psychology, College of Education, at the University of Illinois at Urbana–Champaign. He is the founder of internship and fellowship programs to support graduate and postgraduate students of color from traditionally underrepresented communities (e.g., the Graduate Education Diversity Internship Program of the American Evaluation Association) and served as the 2012 president of the American Evaluation Association.

Ernest R. House, PhD, is Professor Emeritus at the University of Colorado Boulder. Previously, he was a professor of education at the Center for Instructional Research and Curriculum Evaluation at the University of Illinois at Urbana–Champaign. He has been a visiting scholar at the University of California, Los Angeles; Harvard University; University of New Mexico; Center for Advanced Study in the Behavioral Sciences, Stanford University; and universities in England, Australia, Spain, Sweden, Austria, and Chile.

Alana R. Kinarsky, PhD, is a postdoctoral scholar at the University of California, Los Angeles, School of Education and Information Studies, specializing in social research methodology with a focus on evaluation policy. She also works as an independent evaluator in Portland, Oregon, and chairs the American Evaluation Association’s Evaluation Policy Topical Interest Group.

Jean A. King, PhD, is Distinguished Teaching Professor Emerita in the Department of Organizational Leadership, Policy, and Development at the University of Minnesota. She founded the Minnesota Evaluation Studies Institute in 1996 and served as its director for 23 years.

Melvin M. Mark, PhD, is Professor of Psychology at The Pennsylvania State University. He has served as president of the American Evaluation Association and as editor of the *American Journal of Evaluation* (for which he is now Editor Emeritus).

Donna M. Mertens, PhD, Professor Emeritus at Gallaudet University, has over 50 years’ experience in evaluation. She served as president of the American Evaluation Association.

Michael Quinn Patton, PhD, is the founder and director of utilization-focused evaluation and is a former president of the American Evaluation Association. A sample of his workshop presentations can be viewed at his YouTube channel at www.youtube.com/channel/UCd7qY-3bQSZV7cUYmmQ_JUmA/videos.

Hallie Preskill, PhD, is an independent evaluation and learning consultant providing thought partnership and coaching to philanthropic and nonprofit leaders and staff. From 2009 to 2021, she was a managing director with FSG, where she oversaw the firm’s Strategic Learning and Evaluation practice and advised on a wide variety of evaluation and learning projects with foundations, nonprofits, corporations, and government organizations.

Michael Scriven, PhD, is Professor of Psychology at Claremont Graduate University and a senior research associate at the Evaluation Center, Western Michigan University. He has served as president of the American Educational Research Association and the American Evaluation Association.

Vidhya Shanker, PhD, is an interdependent evaluation scholar, practitioner, activist, organizer, and field-builder. She convened The May 13 Group, which includes initiatives like *The Power of Perspective: Generations of Evaluators Generating Change*, *A People’s History of Evaluation*, and *Repairing (and Infiltrating) the Canon: An Oral History Project*. In addition to founding the Collective Knowledge Works, LLC, she currently co-chairs the American Evaluation Association’s Diversity, Equity, and Inclusion Working Group.

Robert E. Stake, PhD, was Director of the Center for Instructional Research and Curriculum Evaluation (CIRCE) at the University of Illinois at Urbana–Champaign from 1975 to 2016. Among the evaluative studies he has directed have been works in science and art education, model programs and conventional teaching, and special education and gender equity.

Naomi A. Stephen, BA, graduated from the University of California, Los Angeles, *summa cum laude*, majoring in International Development Studies. She has worked as a research assistant to Dr. Marvin Alkin for the last few years and is a current PhD student in the Department of Social Research Methodology at UCLA.

Herbert M. Turner, III, PhD, is founder, president, and principal scientist of ANALYTICA, Inc., a woman-owned and minority-led small business. He specializes in using impact evaluation methods to address policy-relevant questions in education and related fields.

Merel Visse, PhD, is Director of the Medical and Health Humanities program at Drew University, Madison, New Jersey, where she was awarded the Meryll Skaggs Award for Excellence in Teaching. She is also Associate Professor in Care Ethics at the University of Humanistic Studies, Utrecht, The Netherlands, where she coordinates the International Care Ethics Research Consortium.