Being an EVALUATOR

Your Practical Guide to Evaluation

Donna R. Podems

THE GUILFORD PRESS
New York London

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

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

Names: Podems, Donna, author.
Title: Being an evaluator : your practical guide to evaluation / Donna R. Podems.
Description: New York : Guilford Press, [2019] | Includes bibliographical references and index.
Identifiers: LCCN 2018021202| ISBN 9781462537808 (pbk.) | ISBN 9781462537815 (hardcover)
Subjects: LCSH: Evaluation research (Social action programs) | Evaluation.
Classification: LCC H62 .P5725 2019 | DDC 001.4—dc23
LC record available at *https://lccn.loc.gov/2018021202*

The comics that appear in Chapters 3 and 6 and the icons were created by Christopher Lysy. Images on the game pieces are from Dreamstime.com, LLC.

I dedicate this book to four very important people in my life:

To my mom, Elizabeth Podems, who, since the first time
I could pick up a pencil, always encouraged me to write

To my dad, Gary Podems, who gave me the courage
to write what I wanted with that pencil

To my children, Gemma and Rhys, who bring amazing joy,
laughter, and fulfillment to my life— and without whom
I would have completed this book 3 years ago!

A Conversation with My Readers

I have written this book just for you: the practicing evaluator; the emerging evaluator; the "I am thinking about being an evaluator" nonevaluator; the "I do monitoring" evaluator; the "I was unceremoniously dumped into evaluation" evaluator; the "I just got out of grad school, and now I have to practice evaluation" evaluator; the "I do not do evaluations [yet]; I just need to know about evaluative thinking" nonevaluator; the researcher; and the person who needs to know how to work with evaluators.

I take each of you on a journey that demystifies evaluation, explores what it means to be an evaluator, and shares some well-kept trade secrets along the way. This is not an evaluation theory book, nor is it a research book. It is a book about being an evaluator who uses evaluation to explore, describe, explain, and eventually judge (in some way) how, and the extent to which, something does or does not work where, for whom, and why. The chapters offer various ways to engage with evaluation information and knowledge—ways that cement learning and encourage reflection. Structured guidance shows you how to untangle various evaluation situations through facilitation, negotiation, and listening. Some core aims of this book are to encourage you to experiment with different ways of thinking; consider multiple perspectives; and acknowledge and engage with the formidable roles that context, power, politics, culture, language, and values play throughout the *entire* evaluative journey.

How the Book Is Organized

I start at the beginning, as one should when writing a book, and then build each chapter on the previous one, providing scaffolding on which to engage in any evaluative process. However, the book does not need to be read sequentially. I have written the book to support your unique evaluation adventure, which may begin in Chapter 3 or Chapter 14, then jump to the last section of Chapter 5, and then circle back to a paragraph in Chapter 2. The book

provides a structure that untangles the messiness of evaluation by informing and guiding the many choices faced, and the various discussions held, in an evaluation journey; at the same time, it allows for a realistic sense of unpredictability and differences found in each one.

The book has two parts. Part I lays the foundation for exploring any evaluative process: It breaks down and takes you through what an evaluator needs to know, whether it is an evaluative process for learning, reflection, improvement, accountability, social justice, or judgment. Part II focuses on working as an evaluator and exploring evaluation: It is aimed at fostering reflection on, and thoughtfulness and awareness of, the many kinds of evaluation that exist; discussing the many evaluative roles that can be filled; and exposing some common challenges and pitfalls often encountered in the field, but rarely described in textbooks. Fused together, Parts I and II demystify evaluation and provide a firm basis for candidly engaging with any monitoring or evaluation process, no matter what role you fulfill.

Throughout the book, I dive into the murky sea of evaluation and guide you through it all. Although I provide strategies and processes, I do not provide a "do this and then do that" model, which may work in some situations, and then not in others. Rather, I offer a way to *think* through any evaluative process so that you can comfortably engage with a peer, colleague, boss, beneficiary, or client in any evaluation situation. The book will support your work in almost any evaluation context—whether you work for a nonprofit, a community-based organization, a donor, a government, a university, an institute, or a foundation, or for some other group that aims to fix, change, influence, or in some way make something better in the social world.

The Learning Approach

People learn in different ways and at different speeds. A concept that seems easy to grasp for you may be a stumbling block for others. This is true in most of life. Some people find it harder than others to tell time by using an analog clock, or instantaneously to know their left from their right. Often there is something in our daily lives that we find more challenging to do or understand than the people around us; often we do not talk about this. The same is true in evaluation. With this recognition, I draw on different ways of learning, including facilitated interactive activities, self-learning exercises, areas for reflection, sections for discussion, and practical applications. Furthermore, we all have lives outside of practicing evaluation. Some days I find I have a few minutes where I can quickly watch or listen to something, and some evenings I find an hour or two to read. Once in a blue moon, I find more focused time. Acknowledging that reality, I provide further ways to learn about each chapter's topic that meet your varying needs—carefully balancing how much more you want to learn with how much time you have to learn more.

Remember Aesop's fable about the Tortoise and the Hare? The overconfident, sleek, fancy-looking Hare bragged about how fast he could run. Tired of hearing him boast, the Tortoise challenged him to a race. The race was long and challenging, and the Hare sprinted

ahead, while the Tortoise approached the race in a precise and methodical way. All the animals in the forest gathered to watch the race, which was won by the wise Tortoise. In the book, we (you and I) are Tortoises. We do not rush to the end to assess and judge; rather, we thoughtfully engage and delight in each step of the evaluative process.

My Perspective

Making sense of things in the social world is not an easy task for anyone. Being asked to judge something, and in doing so to value that thing, can be daunting. There is no perfect template. An evaluator needs to engage in the academic theory and the practical side of evaluation; it is not an either–or decision. Theory informs practice, and practice informs theory. An evaluator educates when appropriate, to ensure engagement, learning, and meaning in the process and in its results. An evaluator comfortably negotiates where necessary, and acknowledges and engages with culture, language, power, politics, and values that constantly swirl in the evaluation process and influence all decisions, including her own. (See "My Use of Pronouns," below.) Through demystifying the process, an evaluator invites people into it and enables them to join in; he actively takes that responsibility, and in doing so through evaluation makes the world a bit better off, even if just a tiny bit.

Why I Wrote the Book

The reason I am an evaluator is not that I enjoy judging things or being critical. I am an evaluator because I believe that evaluators have the potential to make the world a better place to live in. I have written this book to provide a way of thinking that supports an evaluator—not to intimidate others with knowledge, but to engage and guide them through a transparent process.

All over the world, especially where funds dedicated to social improvement are finite, thoughtful, kind, knowledgeable, and skillful evaluators can offer appropriate evaluative processes that inform decisions that affect people's lives, animals' lives, and the environment. Capable evaluators can fulfill an important societal role, while incompetent ones can be a detriment.

Some Guiding Icons

In an evaluation report, it is very common to have a page that lists all the abbreviations used in the report, along with their full terms; this provides a quick guide that you can refer to when reading the report. In this book, I use eight icons to draw your attention to how various features support you in more easily engaging with the text. The icon list follows.

Activities or things to do

Advice

Sharing a trade secret

Suggested resources

Clarifying questions

Practical conversations

The Use of Pronouns

The use of pronouns is a funny thing. In older books, there is a constant use of *he*. Being a *she* myself, I find that a bit bothersome. On the other hand, when I am reading, it is not always a smooth experience to read *he/she*. For some reason, when I read *he/she*, my brain always reacts with "Wait, so which is it? A he or a she?" This makes me stop reading the sentence, decide if it is a he or she, and then reread the sentence with my chosen pronoun. So I decided to alternate between using *he* and *she* throughout the book. But this then got me thinking: "What about people who do not identify with those pronouns? Perhaps I should use the pronoun *they*?" However, the use of *they* as a singular pronoun is not yet that common, and some early reviewers of the book kept thinking I was incorrectly using the plural *they*. So a note to my readers is that I use *he* and *she* alternately throughout the book, sprinkled with a few uses of *they*, to refer to a person.

Furthermore, I often switch in the book to using the term *we* instead of *you*, when you are technically not part of the *we* conversation just yet. I often find that using the term *you* can be heard as too direct, too confrontational, or too "othering." That is, it can be perceived as an attack that puts people on the defensive or makes them feel isolated. Of course, there are times when only *you* is appropriate, or indeed is the only option. However, using the term *we* underscores that we (you and I) are on the evaluation journey together.

Acknowledgments

If someone asked me to describe how this book came about, it would be a very long story. The story starts with my parents, Elizabeth and Gary Podems, who stood behind me and supported me in whatever I decided—whether it was passing up law school to join the Peace Corps, leaving a paying job to take a volunteer fellowship in South Africa, or living in Bosnia or Somaliland, because, as I explained to them, it just seemed right to me. All those journeys, and many more, informed the kind of evaluator I am. Mom inspired my love of writing by giving me my first journal at the age of 9, and then a new one every year after that.

Michael Quinn Patton entered my life at the beginning of my evaluation journey as my teacher, and has continued in the role of mentor and friend. Throughout my career, he has encouraged me not to conform to evaluation norms when they did not resonate with me (or my clients), to stay true to my principles and values, to march to the beat of my own drum (not his or anyone else's), and to write. He gave me the greatest gift: He believed in me.

I would not even be an evaluator had it not been for Paula Bilinsky, who plucked me out of nowhere and put her trust in a newbie just back from the Peace Corps; she patiently explained the basic ropes of evaluation, and has remained a constant and valued friend in my evaluation journey. Jennifer Greene grounded me in qualitative research and introduced me to feminist and democratic evaluation, all of which heavily influence my practice. She then went on to encourage me to edit my first book, which gave me the courage (and experience) to finish this one. After grad school, I approached C. Deborah Laughton and asked her guidance on how to publish a book; she provided a path that I diligently followed over the next 13 years, which proved to be excellent advice. I thank her for continued support during the making of this book.

Over the years, several of my colleagues and clients provided insights found in this book, as well as much-needed encouragement, for which I am most grateful. The book was a whisper of an idea when Catie Lott encouraged me. As it was taking shape, Kerstin Rausch Waddell provided a calming and insightful voice, which I relied on heavily throughout the

journey. Anna Davis never stopped asking me when I was going to write the book, read multiple drafts, and gave me sage advice and thoughtful feedback. I stalled in the middle, and Tessie Catsambas provided a timely nudge. Benita Williams offered courage when I needed it most, along with her quantitative insights and the best-ever pep talks. Susan Tucker provided inspiring weekend and evening conversations that kept me on my toes. A book does not get written overnight (at least this one did not), and during the 5 years of writing, many other people have informed its shape and its content. I thank Leanne Adams, Daleen Botha, Cindy Clap-Wincek, E. Jane Davidson, Svetlana Negroustoueva, Tim Reilly, Gäelle Simon, Liezel de Waal, and Lauren Wildschut.

Robin Miller asked me to fill in for her and teach her evaluation foundation course at Michigan State University when I was a few years into writing the book. Teaching that class provided new insights, and Robin's thoughtful feedback on the second part of the book inspired me to think harder. Aimee N. White was supposed to be a "blind" peer reviewer; however, she knows my writing and stories so well that she guessed it was me. Her funny and sincere support, and meticulous feedback in the final, crucial stretch to the book's completion, were invaluable. I also thank the many peer reviewers who did not recognize me, and whose thoughtful criticism made this a better book: Mary E. Arnold, College of Public Health and Human Sciences, Oregon State University; Penny Burge, School of Education, Virginia Tech; Janice Fournillier, Department of Educational Policy, Georgia State University; John Klemanski, Department of Political Science, Oakland University; and Neil J. Vincent, Department of Social Work, DePaul University. Among the important elements of my book are its icons, comics, and game pieces, and I thank Christopher Lysy for enhancing the book with these fun and "totally me" additions.

Two very important people during the whole process were my children, Gemma and Rhys. I thank them for giving me the peace and quiet time to write (OK, I admit, it was often in exchange for iPad time) and for always asking, "How's the writing going, Mom?" Thank you, my Munchkins.

Contents

PART I DOING EVALUATION AND THINKING EVALUATIVELY 1

CHAPTER 1 SPEAKING THE LANGUAGE 3

The Power of Words 4
Why Evaluation Terms and Concepts Can Be Perplexing 6
Why an Evaluator Should Clarify Terminology 6
The Evaluator's Role in Clarifying Terminology 7
Learn the Meaning—Not Just the Label 9
Evaluation, Evaluative Thinking, and Monitoring: What These Terms (Can) Mean 11
Intervention, Activity, Project, Program: Doppelgangers or Differences? 25
Vision, Mission, Goal, Objective: Management Speak Translated into M&E Speak 25
Wrapping Up 26
Our Conversation: Between You and Me 27

CHAPTER 2 THE TALE OF THE RESEARCHER AND THE EVALUATOR 29

Why It Is Important to Understand the Differences between Researchers and Evaluators 30
Researchers and Evaluators: Compatriots or Competitors? 30
How to Facilitate Useful Conversations 32
Our Conversation: Between You and Me 35
Describing the Evaluative Journey 36
Potential Users of the Evaluative Journey 37
Wrapping Up 38

CHAPTER 3 STARTING THE EVALUATIVE JOURNEY 39

Introducing the Intervention 40
Introducing the Problem Statement 44
Clarifying Different Kinds of Problems, and Why It Matters 51
Introducing Beneficiaries and Six Related Terms 50
Focusing on the Intended Beneficiaries 51
Including Beneficiaries in the Process: Placebo or Panacea? 56
Wrapping Up 59
Our Conversation: Between You and Me 59

CHAPTER 4 How We Know What We Know, and Why We Think That 70

The Conundrum 70
The Academic Discussion 71
The Practitioner Conversation 73
The Difference between Facts and Assumptions 75
Engaging with Facts: Before and After the Fact 77
How Mixing Facts and Assumptions Can Influence an Intervention 82
Values, Values Everywhere 84
Wrapping Up 86
Our Conversation: Between You and Me 86
Postscript: Facts and Assumptions in Dickens's "A Christmas Carol" 87

CHAPTER 5 Data and Credibility: What Inquiring Evaluator Minds Need to Know 88

Qualitative Inquiry 88
Quantitative Inquiry 94
Mixed Methods 99
Sampling 100
Practical Guidance in Choosing an Inquiry Method 102
Credible Data 103
Credible Evidence 104
Baseline Study 106
Wrapping Up 109
Interlude 110

CHAPTER 6 Linking Problem Statements, Interventions, and Results 112

The Importance of a Plausible and Evaluable Intervention 112
What Needs to Be Linked 115
How to Identify Strong Links and Broken Ones, and Why That Matters 116
Wrapping Up 123
Our Conversation: Between You and Me 123

CHAPTER 7 All about Results 124

Four Critical Reasons to Understand the Logic of Results 125
Starting the Results Discussion 126
Logical Thinking in the Results Discussion 127
An Evaluator's Role in the Results Discussion 128
Why Discussing the Logic of Results Is (Often) Demanding and Messy 128
Strategies for Unpacking the Logic of Results: Mastering the Results Conversation 130
The Framework: Unpacking Results 132
Contextual Factors: Messing with the Results Logic 145
Wrapping Up 149
Our Conversation: Between You and Me 150

CHAPTER 8 Talking Intervention Theory (and Logic) 151

Three Reasons to Have an Explicit Theory of Change 152
Four Examples of How an Evaluator Engages with a Theory of Change 153
Sorting Out an Intervention's Theory of Change: A Framework 154
Common Challenges to Discussing Theory of Change 156
Program Logic and Theory of Change: How They Are Connected, and Why Both Are Useful 159
Three Things We Know When Theory of Change and Logic Are Clear 162

Theory and Logic: Informing When to Monitor and Evaluate What and Where 163
Wrapping Up 171
Our Conversation: Between You and Me 172

CHAPTER 9 ASSESSING AND EVALUATING PROGRESS 174

Measurement and Assessment: The Differences 175
What to Assess 176
How to Measure Progress 181
Wrapping Up 196
Our Conversation: Between You and Me 197

CHAPTER 10 THINKING THROUGH AN EVALUABLE INTERVENTION: PULLING IT ALL TOGETHER 198

The Baby Game 199
Common Challenges with M&E Frameworks 203
Wrapping Up 206
Our Conversation: Between You and Me 207

PART II WORKING AS AN EVALUATOR AND EXPLORING EVALUATION 209

CHAPTER 11 THE PERSONAL CHOICES OF BEING AN EVALUATOR 213

What Type of Evaluator Do You Want to Be? 213
Being an Evaluator: An Abundance of Roles 213
An Evaluator's Role by Evaluation Purpose: Formative, Summative, and Developmental 217
An Evaluator's Role by Evaluation Purpose: Four More of Them? 220
An Evaluator's Role through Self-Identification: Sector Specialist or Evaluation Generalist? 223
An Evaluator's Role through a Formal Position: Internal or External Evaluator? 224
An Evaluator's Role in Some Not-So-Common Positions 227
Exploring Evaluation's and an Evaluator's Roles in Society 228
Who Is Considered an Evaluator, and Who Is Not? 233
Wrapping Up 235
Our Conversation: Between You and Me 236

CHAPTER 12 THINKING ABOUT VALUES 238

What Are Values? 239
Ethics, Values, and Morals: The Practical Differences 239
Exploring Your Values 239
Where to Find Values to Inform the Evaluative Process 240
The Evaluator's Role in Identifying and Applying Criteria to Value an Intervention 244
Evaluation Ethics and Principles 251
Wrapping Up 255
Our Conversation: Between You and Me 255

CHAPTER 13 THINKING ABOUT POWER, POLITICS, CULTURE, LANGUAGE, AND CONTEXT 257

Evaluation Choices: A Little Bit Technical 257
Evaluation Influencers: The Fabulous Five 258
Evaluation Context: Broadest of Them All 275
Wrapping Up 277
Our Conversation: Between You and Me 277

CHAPTER 14 THE SCHOLARLY SIDE OF BEING AN EVALUATOR 278

Knowing the Evaluation Literature: It Matters 278
Ways to Start Engaging with the Literature 280
Wrapping Up 283
Our Conversation: Between You and Me 284

CHAPTER 15 NAVIGATING THE MAZE OF EVALUATION CHOICES 285

When Someone Wants an Evaluation: The Names They Use 285
Implementation, Outcome, and Impact Labels 286
Approaches to Evaluation 296
The Evaluation Design 310
The Evaluation Report and Other Ways to Communicate 311
Evaluating the Evaluation 314
Wrapping Up 315
Our Conversation: Between You and Me 316

CHAPTER 16 THE WORLD OF RECOMMENDATIONS (AND THE UNDERWORLD OF CRITICAL FEEDBACK) 317

A Recommendation: What Is It? 317
Who Are the Recommendations For? 319
The Evaluator's Role in Recommendations 320
Strategies for Developing Meaningful Recommendations 320
Strategies for Presenting Acceptable Recommendations 321
A Changing World: Shifting from Recommendations at the End to Learning Throughout 324
Providing Feedback 325
Wrapping Up 327
Our Conversation: Between You and Me 327

CHAPTER 17 THE DIRTY LAUNDRY CHAPTER 329

Eight Challenges in the Design Phase, and How to Mitigate Them 330
Ten Challenges to Implementation and Reporting, and How to Mitigate Them 335
Nine Common Questions That Arise during an Evaluative Process 342
Wrapping Up 346

EPILOGUE 347

REFERENCES 349

AUTHOR INDEX 356

SUBJECT INDEX 358

ABOUT THE AUTHOR 368

Purchasers can access an online supplement at *www.guilford.com/podems-materials* that contains some of the activities and games described in the book.

PART I

DOING EVALUATION AND THINKING EVALUATIVELY

An evaluator should understand how to gather data and use that to make a finding, a value judgment, and then develop a recommendation. I think being able to sustain that type of clearly articulated thinking throughout the conceptual pathway is a rare talent. . . .
—DUGAN FRASER, Chairperson of the South African M&E Association (2015)

CHAPTER 1
Speaking the Language

An evaluator is a human being living in a complex world. She (or he or they) engages with other human beings through a process that values something that is important to someone in some way. That process is called the **evaluative process,** and its findings are used to learn, improve, judge, or in some way inform decisions. How an evaluator thinks, who and what shapes her thinking, and how she views evaluation's role in society are all germane to how she defines herself as an evaluator. (For my use and choice of pronouns, please see page x in "A Conversation with My Readers" at the start of the book.)

Consider that the multidisciplinary field of evaluation includes pluralist philosophies that draw on multiple methods, criteria, measures, assessments, perspectives, audiences, and interests, as well as a plethora of traditions that have emerged from different social science disciplines (e.g., anthropology, psychology, sociology). Add to all this the roles of politics, power, culture, language, and values, and you can begin to see the many difficulties of definitively explaining what an evaluation is and how to design and implement one. Nonetheless, definitions and guidance abound, although of course they do not always agree.

Mix into this conversation the many roles that an evaluator can play, and you begin to see the complications of defining who is an evaluator, or drawing a boundary around what an evaluator does. Since the late 1980s, evaluation associations and societies, governments, international aid organizations, foundations, institutes, and other groups and institutions have been exploring ways to define who is an evaluator. However, evaluation differs from the medical, accounting, or legal fields, for example, in that no formal certification process is needed to practice evaluation: Anyone can claim to be an evaluator (McKegg, 2017; Podems & King, 2014).

Defining and clarifying what an evaluator does, how to be an evaluator, and the essence of being one constitute this book's core focus. Being comfortable with discussing the common language of evaluation provides a starting point from which to demystify evaluation and explore how to be an evaluator. If you would like to skip ahead to an in-depth discussion of what it means to be an evaluator, read Chapters 2 and 11. Otherwise, join me at the start of our evaluation journey, with a discussion of evaluation language.

The Power of Words

I begin with a memory of an awkward moment.

> "I grew up in a small town in New Jersey, and moved to South Africa in 1998. Back in those early days, I did not know too many people in South Africa. As the December holidays drew near, I received several invitations to holiday parties, and was especially glad to accept an invitation to join my neighbors for Christmas dinner. I was raised never to go to someone's house empty-handed, so I immediately asked what I could bring to the meal. My neighbor asked me to bring some crackers. I was glad for such an easy request. So I showed up on Christmas Day with two boxes of crackers; one salted and one flavored with rosemary. My hostess graciously accepted them. As we sat down to dinner, her husband loudly asked, 'Where are the crackers?' Since they were on the table, I pointed to them, at which point everyone burst out laughing. Apparently crackers are also small, prettily wrapped pieces of round cardboard, each containing a small prize; when you pull one open, it makes a small pop, like a firecracker. Who would have thought? I mean, who stops to ask, 'Tell me, what do you mean by a cracker?'"

Words may have different meanings to different people, in different contexts, and points in time. Sometimes the same word simply has multiple meanings for the same person. How and why people choose the words they use, and how their meanings are interpreted, are all part of everyday life; they are also relevant to any evaluative process. A critical part of being an evaluator is speaking the language. To do this, an evaluator needs to know the difference between evaluation words, terms, and concepts with concrete definitions that an evaluator should confidently know and be able to explain, such as *random sampling* (see Chapter 5), and words with meanings that vary by user and context, where evaluators often need to negotiate or clarify definitions and practical applications, such as the term *impact.* Although throughout the book I provide concrete definitions for evaluation terms and concepts that have them, the present chapter specifically looks at words and concepts with meanings and practical applications that need to be clarified prior to engaging in any evaluative process.

Language is a powerful tool, and a separate, broader conversation (which builds on the one found in this chapter) on how language influences evaluation is provided in Chapter 13. We start the evaluative journey here by exploring words, terms, and concepts that often perplex through their various definitions, interpretations, and applied meanings, and that have the potential to derail any evaluative process if they are not properly considered.

Before delving into examples for the evaluation field, let's start with some from everyday life. What would you think if someone said that the movie they just watched was "bad" or "sick"? Depending on the tone, the expression, the age of the speaker, and the cultural context, it would likely have different meanings, from "The movie was awful" to "It was fabulous." And because words have these nuances, there are always times in our lives where we feel we are the only ones in the room (or the world) who do not understand a term, or completely misunderstand it. Recently a younger colleague of mine asked what I was doing for the weekend. I responded that I was working on this book. She muttered something that sounded like "FOMO." I laughed and she laughed, but I was thinking, "What in the world is FOMO?" For those who do not know, it means Fear Of Missing Out. I now know this, but I didn't know it then.

Consider a time when someone used a word or mentioned a concept that you thought you should know but did not, or you thought everyone else in the room knew and you did not. Most likely you nodded and pretended that you understood, and then did a web search on it later (or perhaps in the moment, if you had a chance to do so surreptitiously). Nearly everyone has been in this situation when it comes to working in the evaluation field. Ironically, it almost always starts with the mere mention of the field's most fundamental concept: evaluation.

The term **evaluation**; its cousin, **monitoring**; and the phrase and acronym that link them together, **monitoring and evaluation (M&E),** are among some of the more problematic terms in the field. While some definitions differ only slightly, others blatantly disregard and sometimes flat-out disagree with each other. These factors alone can create a breeding ground for intimidation (and perhaps insecurity, frustration, exasperation, or confusion, to name a few descriptors mentioned by colleagues and students) in the evaluation field. As one of my peers, Anna Davis, once told me, "In so many situations, I am sure I know what I am talking about with regard to monitoring or evaluation. Then someone comes along who has more power than me, uses a term in a different way, and tells me otherwise. . . . It is bullying by M&E!"

Thus an evaluator, or any person in a position of power, can indeed be a bully (either knowingly or unknowingly) or can be perceived as one when flinging about evaluation concepts or terms. Or the evaluator can choose to engage in a process that leads to a common understanding of a term. This approach lays the foundation for an evaluation process that is likely to be more useful to all involved.

BOX 1.1. Evaluative Process

Throughout Part 1, the term **evaluative process** is used to describe anything in an evaluation, from conceptualizing the intervention through to the judgment of it. Essentially, the evaluative process can be any part of a process where questions can be unearthed or answers can be found to address the "what," "when," "why," "where," "how," and "for whom" questions.

Why Evaluation Terms and Concepts Can Be Perplexing

Stand a little bit closer to me; I am going to share a trade secret. Evaluation terms and concepts abound, and for many, so do their definitions and their applications. More challenging is that sometimes differences in definitions are not always clear; the differences may lie in the nuances or the unspoken interpretation (as demonstrated through its application) of a definition in a specific setting. There are a few commonplace reasons for why this happens.

- Evaluation has emerged from diverse social science disciplines (e.g., anthropology, psychology, sociology), and these disciplines themselves use words and terms differently.
- Groups or individuals in positions of power, such as governments, donors, institutes, and foundations, are emboldened to define evaluation terms and concepts, and do so within their own worldviews. Their power stems from various sources; the two I mention here, money and influence, are tightly connected. Briefly, those who hold the purse strings can often define and apply words in any way they wish. If a program or intervention wants to receive funds, for example, more than likely it is the donor's definition that is used. This is often true even if that definition varies from other, more common understandings and uses.
- The innumerable evaluation textbooks, websites, evaluation guides, and blogs, which offer varying definitions of the same term, do not escape blame.

Why an Evaluator Should Clarify Terminology

Engaging explicitly with others about the meanings of evaluation terms and concepts *before* the start of any evaluative task will lay the foundation needed for a useful evaluation. These conversations help to avert misunderstandings, frustrations, potential conflicts, and possibly grave disappointment at the end. Furthermore, clarifying the terminology at the start of an evaluation process is a neutralizing activity. The start of an evaluative process is often nerve-wracking for those being evaluated, and creating the space to clarify terminology provides a neutral, nonjudgmental place in which to calm nerves and encourage dialogue and transparency.

Never assume that people understand evaluation terms or concepts, what they look like in practice, or how they influence an evaluation. When you are clarifying evaluation terminology, separate the terms into two groupings (Patton, 2012). The first group should contain words and concepts with concrete (though often misunderstood) definitions. Address these first. The first step for each of these terms is to provide a clear definition. The second step, and where the discussion is concentrated, is to describe how that term will be applied and how that application will influence the evaluation. An example of a word with a specific definition that commonly needs a thorough discussion is the term *causality*. The term has a concrete definition; there is no negotiation along the lines of "I think it means this, and he thinks it means that"; it means what it means. How it is applied in the evaluation is where the beneficial discussion happens. Do not assume that knowing the concrete definition automatically leads to an easy conversation; the conversation may still bring a challenging interaction. To read a technical discussion on causality, please skip ahead to Chapter 15.

The remainder of this chapter focuses on the second group: the words and concepts that have fuzzy and varying definitions. In each of these instances, discussing and negotiating the definition are wrapped up in the discussion of how it will be applied, making for an often complicated yet much-needed conversation. Regardless of the group into which a word falls, each conversation will, at the very least, support the evaluation's transparency, credibility, and use; *at the very most,* it will contribute to an engaging, collaborative, and social equity–building process.

The Evaluator's Role in Clarifying Terminology

When terms have multiple definitions and/or meanings, evaluators can often find themselves in the position of needing to clarify the terminology. As such, an evaluator can serve in three roles: as an educator, negotiator, *and* facilitator, so that each concept and its application are agreed upon for that specific evaluation. An intimate knowledge of the multiple definitions for the same term, and the subtler nuances that exist, will allow the evaluator to initiate and facilitate fruitful discussions. In turn, these discussions support equitable engagement, transparency, and clear expectations among those involved in the evaluative process of how the term will be applied, and how that will influence the evaluation. As suggested above, this process ideally increases the likelihood of a useful evaluation.

Throughout this book, the evaluator is positioned as one who leads processes to untangle meanings, clarify concepts, and concretize the application of evaluation terms, which supports evaluation use. In this way, I unashamedly lay the responsibility of a clear evaluation process in the hands of the evaluator. Do not interpret this to mean that the evaluator should hold all the power in an evaluative process; rather, the evaluator recognizes who holds power, acknowledges how and why that power is held, and engages with terminology

so that power is more evenly distributed. After all, language is a major form of power. For a more specific discussion of language and power in an evaluation, skip ahead to Chapter 13.

BOX 1.2. The Evaluator's Role in Clarifying Terminology

The evaluator's role is to comfortably engage with terminology and, whenever needed, to clarify it with evaluation users and those involved in the evaluation process. To do so, she needs an in-depth knowledge of evaluation, as well as facilitation, negotiation, and education skills.

I am an evaluator for a small organization. I find that sometimes evaluation words are not concrete, and I need to negotiate their meaning with the donor who provides the funding. What if the donor who provides the funding disagrees with what my internal evaluation guide says?

The uses of words that do not have concrete definitions are often determined by those who have the most power. Commonly, though not always, these individuals or groups are those who provide or control the resources or make other key program decisions. If a donor or government agency that provides the bulk of the intervention's funds (and/or is your boss) insists, "This is how we define X," this is likely not a good situation in which to have an argument premised on "My definition is right and yours is wrong, and here is the guidebook that backs this up." Rather, what is key is that there is a shared, explicit understanding of how the term will be defined and applied in the evaluation. I strongly suggest that this be written, not just orally agreed upon, and that all people involved agree that *this* is indeed the agreed understanding. The most worrisome thing about providing advice in a book is that there will always be exceptions to the rule; there may be times when bringing your definition, or a different definition, to the discussion is useful. Ask yourself, "How will having this conversation, clarifying this term, and using that definition contribute to a useful evaluation?" And ask the opposite: "If we use that definition, how could it create challenges, and to whom?"

There may be one factor that makes everything a bit tricky: the personal factor. Even if the funding source or management structure does not change (i.e., the government department or donor remains the same), the discussion of terms may still need to be revisited throughout the life of the program or the evaluation. When a new *person* fills the decision-making or funding role, they may bring new interpretations to evaluation terms and processes (often without explicitly stating this), which will leave the recipients confused and often frustrated. The personal factor influences all evaluation processes.

This last point reminds me of another trade secret I must share with you: The personal factor is one of the biggest undiscussed influences in any (and all) evaluative processes. Recognizing,

and not ignoring, the fact that individuals influence an evaluation process is critical, as each person can be a tremendous facilitator, an impediment, or just a source of constant bumps and bruises along the way.

I am an evaluator for multiple interventions in a large nonprofit organization. As such, I have multiple donors who all bring different understandings to the same evaluation terms. How do I straighten this all out?

Working in an organization that is accountable to several different donors, departments, or managers, all of whom use the same evaluation terms or concepts in dissimilar ways (or sometimes with just enough nuanced differences), can cause havoc. I recommend starting by clarifying *who* means *what* for each evaluation concept or term, and how each meaning translates into practical application (even for the words with concrete definitions). Then determine in what way practically applying the different definitions and/or interpretations of the same evaluation word influences how to conduct an evaluative process. Compare how the different interpretations will influence what is done; will they look the same in practice? If not, clarify how the differences influence how the intervention will be evaluated and valued, and use that information to have an informed discussion on how to move forward.

LEARN THE MEANING—NOT JUST THE LABEL

Evaluators need to use a term or concept with the same understanding as their peers, colleagues, or clients to ensure a useful evaluation. It is with the utmost hesitation that I use the following example to support a conversation about the potential dangers of using labels without having a clear definition of what the terms mean. (Chapter 7 talks about effectively using results labels and refers to the example that follows.) The example is provided to start *our* conversation on labels; it is not a conversation to be used with anyone else on how to use and apply the terms.

Some common labels in evaluation are *input, activity, output, outcome,* and *impact*. Here is one example of how these terms may be used. Some of us monitor how much money is spent to get something (e.g., $100 for running shoes). In evaluation, we call that input, or what was put in to get something out. What the person is doing (e.g., running) is called the *activity*. The very first thing that happens because of that activity, such as how far or how often the person runs every week, is often called the *output*. Some of us may also then monitor how much weight the person loses because he runs so much, which is often what evaluators call the desired or intended *outcome,* or what happens because of the output (e.g., the distances ran). Then, because the person has lost weight, he has higher self-esteem and better overall health, which some evaluators call a *higher-level outcome* (and some call

impact). He then decides to start a new career and adopts an overall healthy lifestyle, which brings him greater happiness, and some refer to this result as the *impact* (and some call it *longer-term impact*).

This example illustrates a common, simple, and quick way of using these labels without much concern for the dangers of doing so.

I now implore you to be patient and trust our journey. I ask you not to argue with or question what is written in the previous paragraphs (not just yet, anyway), and to not immediately use these labels mechanically as described. Rather, I invite you first to read the next section about the dangers of using labels in evaluation. If you want to jump ahead to read an in-depth discussion on sorting out the input–output–outcome–impact (and so on) discussion, go to Chapter 7 (though I encourage you to keep reading from here through Chapter 10 to see all the linked discussions). After reading *through* to Chapter 10 (or, if you must, after only reading Chapter 7 on results labels), revisit the example that starts on the previous page. You will then understand why I have asked you not to argue (just yet) or to use the labels without a deeper understanding of the questions they pose. I'll see you on the other side.

Learn the Meaning—Not Just the Label

Labels can be useful when we buy a product. For example, "Here is shampoo for babies, and there is shampoo for dogs." Or "That is dog food, and this is cat food." Or "That is poisonous." I buy cat food for my cat, and dog food for my dog. Labels are meant to be helpful. To be honest, however, my dog seems to prefer the cat food, which she always steals from the cat; as such, this provides my first practical example of how perhaps labels are not always helpful in the way we think they should be. Although my dog's eating cat food is expensive for me (as she steals copious amounts), there is little harm done from her perspective (but from the cat's perspective, he may disagree!). It would be much worse if my dog ate something labeled "poison." In M&E, some labels can be just as wily.

Do not get caught up in label-fueled arguments, such as "That is monitoring!"/"No, that is evaluation!" or "That is an impact!"/"No, it is an outcome!" These discussions will exhaust you. If someone uses an evaluative term (even when the word or concept has a concrete definition), ask them what they mean by it. *Descriptions* under the labels are important. When you are engaging with a client, colleague, or friend about an evaluation process or term, the primary goal is for everyone to have the same understanding insofar as what is needed, intended, and expected. Use a label when, and only when, the meaning is explicitly clear to everyone.

I have now talked about the importance of knowing when a word has a *concrete* definition and when it has a *fuzzy* one; about the need to be wary when using evaluation labels; and about the role of the evaluator in bringing clarity to each context. Now let's apply what we have just learned by discussing three convoluted terms in the field.

Evaluation, Evaluative Thinking, and Monitoring: What These Terms (Can) Mean

The terms **evaluation, evaluative thinking,** and **monitoring** are often used together, and sometimes interchangeably, in an evaluative process. As the next several pages illustrate, there are many "correct" definitions and understandings for each term. However, for many of these words, there is also a non-negotiable core element (or set of elements) in the definition. These elements can be described in multiple ways by various people, implicitly or explicitly, which adds to the trickiness.

Together, let's explore these three terms, and explore how to discuss and engage with them so that their meanings are clear and support straightforward evaluative processes. Facilitation and training tips are provided that can be used to promote self-reflection or to engage others in a discussion aimed at creating understanding and agreement, or simply at clarifying disagreement (the old "We agree to disagree" sentiment). Each is useful in its own way. Let's start with the word *evaluation*.

Evaluation

What is *evaluation*? The word has different meanings and associations for different people, who are informed by, among other things, their education, cultures, values, and exposure to various evaluation experiences. Add to this that organizations, evaluation theorists, social science disciplines, governments, foundations, institutes, and donors often have their own slightly nuanced definitions, and sometimes what appear to be just completely different understandings. Thus asking more than one person, "What is evaluation?" rarely results in a specific and consistent answer. An even more complicated discussion revolves around who is an evaluator, and who is not, as discussed in Chapter 11 of the book. For now, let's focus on uncomplicating the term *evaluation*.

Thomas Schwandt (2015) offers two definitions that provide one place to start a discussion. Schwandt provides a narrow definition of evaluation as a process that draws on research methods to gather information about how a program or policy is working and whether it is effective. He also provides a broader explanation by stating that evaluation is a form of critical thinking that employs evaluation-specific methodologies to judge something. That "something" is broad. I like and will use Schwandt's basic understanding of "something," and I will add a few more throughout this book, some of which are also used by Mathison (2005). When I talk about evaluating "something," I am referring to evaluating an activity, event, intervention, performance, process, product, policy, practice, project, or program. Wait, I am not done yet. Evaluation can also be used to explore and evaluate strategy, systems change, ecological sustainability, resilience, networks, and principles, to give a few additional nouns (Patton, 2015; Rogers, 2016). From this point onward in this text, all of these common nouns will not be written out; this would make the sentences too long. Therefore, when

examples are provided, just one of these words is chosen, most often the word *intervention*. Keep in mind from now on that in any example in any part of the book where the word *intervention* is used, any one of these other common nouns can be substituted.

Two further definitions provided by two influential evaluation theorists, Michael Scriven and Michael Patton, involve different perspectives on evaluation, and therefore the definitions differ slightly. According to Scriven (1991), evaluation is the systematic determination of the merit, worth, or value of something, or the product of that process.

BOX 1.3. Merit, Worth, and Significance

Merit refers to the quality of an intervention—the inherent value. *Worth* refers to how others view the intervention, and determining it involves looking at an intervention in its context. For instance, an evaluator has merit if she brings strong evaluation knowledge and solid experience to an evaluation process. She has worth if those involved in the process have found her knowledge and her services useful (Davidson, 2005; Mathison, 2005; Scriven, 1991). It has *significance* if someone considers what the evaluator did to be special or to have particular meaning.

Michael Patton (2008) offers a broader, more multifarious definition, and one that brings in the idea of evaluation's having multiple purposes, uses, and users. He suggests that program evaluation focuses on three elements: "(1) the systematic collection of information about (2) a potentially broad range of issues on which evaluation might focus (3) for a variety of possible judgments and uses" (p. 39). Patton's definition suggests that different people may have different needs for evaluation. These differing needs are likely to result in the use of different questions, approaches, and criteria to value the intervention.

All of the definitions described in this section are widely acknowledged and offer commonly accepted explanations for the role and purpose of evaluation. However, Schwandt's, Scriven's, and Patton's definitions all stem from different perspectives, leading to some nuanced and some blatant interpretations of the term *evaluation*. None of the definitions provided (Schwandt's, Patton's, or Scriven's) contradicts the others per se, but they give very different understandings of evaluation's role in society, and these differences undergird the discussion about what an evaluator does. Nonetheless, at the core of each definition of evaluation, there are two inherent commonalities.

1. *Evaluation is systematic.* Evaluation draws on systematically gathered evidence. Any evaluation process must have clear, transparent, explicit approaches for collecting and analyzing data. The processes of collecting, analyzing, and interpreting the data must be logically consistent (Patton, 2015; Rossi, Lipsey, & Freeman, 2004; Schwandt, 2015).

2. *Evaluation values or in some way judges something.* The word *value* is part of *evaluation.* (Well, sort of, as we do need to add an *e*.) Nonetheless, an evaluation provides more than just findings; it gives a value judgment. Here is where one of the most complicated parts of evaluation often comes in; after all, a value is based on assumptions or worldviews, which are culturally and socially embedded, and very much otherwise context-dependent (Mertens & Wilson, 2012; Patton, 2015; Schwandt, 2015).

There is one more item that is not explicitly stated in these definitions, though it is a core part of defining and discussing evaluation.

3. *Evaluation is political.* A third core element of evaluation is that evaluation is political (Candel, 2018; Chelimsky, 1987; Greene, 2000; House & Howe, 1999; Patton, 2015; Weiss, 1987). What does that mean? Essentially, it means that evaluation can influence—and is influenced by—who gets what, who does not get what, who benefits, who does not, who is even involved, and who is not. Politics are involved in every part of an evaluative process, from the decision to have an evaluation, to how data are interpreted, to how and with whom the knowledge generated from an evaluation is shared, to how decisions are made (Bowman, & Dodge-Francis, 2018; Patton, 2008). See Chapter 13 for a more in-depth look at politics in evaluation.

Hold on a second. Before you continue, I have a question. The word ***data*** *is used a lot in the explanation. What's the difference between* ***data*** *and* ***information****?*

Data and *information* are words that are sometimes used interchangeably, yet they are not the same thing. Data are the raw facts, figures, and words that, when analyzed to make some sort of sense, become information. Information is shaped or interpreted. In academic language, the word *data* is always considered to be plural. For some reason, this is only true in academic writing. In newspapers, for example, it is acceptable to say "Data is." The word *information* is always treated as singular.

Here is an example of data versus information. My son, Rhys, is 9 years old at the time I am writing this. Some days when I go into his room, he has unceremoniously dumped hundreds (dare I say thousands?) of little colored blocks on the floor. To me, it looks like one big mess. To him, it looks like lots of potential items (data) with which to create (interpret) something. And he does. He creates cars, castles, houses, monsters, and more. He takes what looks like a big mess to me (his data) and creates meaning (information) from it, which he gives to me (his client). To state this another way, when data (the individual blocks) are formed into something through analysis and interpretation (my son's creative design process), they become information.

*Given how you have defined **data** and **information**, then what is **evidence**?*

Evidence supports a statement. So data become evidence when used to support an argument, a hypothesis, or a belief. Evidence comes from data, which most often come from one of our five senses—we can hear it, touch it, smell it, taste it, or see it. And other people can, too. I say "most often," because there are other ways of knowing that do not use these five senses (see Chapter 4 for further explanation). Another way to think about evidence is how it is used in a murder mystery. A detective identifies the murderer with her evidence; the evidence consists of the knife, the body, and the killer's motive.

Because of our acceptance of using some words interchangeably in our daily lives, the boundaries among terms such as *data, information,* and *evidence* can be blurry. Discuss the meaning of *data, information,* and *evidence* (and other similar words, such as *knowledge*) with all parties you are working with or for, and ensure that everyone has a clear and shared understanding of their meaning.

BOX 1.4. Core Elements of Evaluation

There are three fundamental answers to the question "What is evaluation?" Evaluation provides systematic, transparent, and logically collected evidence; it offers a value judgment; and it is political.

Now that we are clear on what data, information, and evidence are, and have looked at some basic definitions of evaluation, let's engage in an in-depth discussion.

Where Evaluation Definitions Begin to Differ

The core understanding we have just discussed is that evaluation is a systematic, transparent process; it values something; and it is political. Even when we agree that these are the core elements, however, differences emerge. These differences are often rooted in who provides the definition (e.g., a person's training and experiences), how evaluation is used in an organization (e.g., its formal definition and practical applications), and its perceived role in society. Thus discussing what is evaluation goes beyond providing a rote definition of evaluation to a discussion that encompasses political, social, and cultural worldviews, among others. Remember, when engaging with colleagues or clients about evaluation, facilitate a conversation that leads to an explicit understanding of what evaluation means to them, what they expect evaluation to do, and what they expect to happen because an evaluation takes place. Do not just ask for their definitions (and do not just provide one).

BOX 1.5. The Meaning of *Empirical*

The word *empirical* is commonplace in evaluation. But what does it mean, exactly? The *Oxford Living Dictionaries: English* (2018) website defines *empirical* as follows: "Based on, concerned with, or verifiable by observation or experience rather than theory or pure logic." I would supplement that definition with the statement that the means of gaining that knowledge need to be transparent and logically consistent.

To encourage learning and dialogue, activities are suggested throughout the book. The first activity described on the next page addresses the term *evaluation*. All activities draw on adult learning principles, which assert that adults' knowledge and life experience should be recognized and drawn upon in any learning opportunity. Drawing on adults' previous experience with conducting, managing, or observing evaluation through their work, or even in their daily lives, will help to ensure that conversations resonate with them, which in turn helps to ground learning. To conduct the activities, facilitation skills are critical. If you are still building those, consider working alongside a skilled facilitator who can demonstrate these skills, and from whom you can learn (i.e., a knowledgeable other).

BOX 1.6. Knowledgeable Others

As we move through the book, you will realize the many skill sets and knowledge areas that an evaluative process may require. When building knowledge and skills as an evaluator, it is often useful to draw on *knowledgeable others*. Susan Tucker, an education evaluation specialist, describes knowledgeable others as people who bring in knowledge and skills that supplement your own knowledge and skill set. These may be knowledge and skills that you have not yet acquired, or never plan on acquiring due to lack of time, differing interests, or other factors. Evaluators often (though not always) work in teams that include knowledgeable others, or subsequently draw in knowledgeable others as needed. In any evaluation process, including knowledgeable others often creates a richer, more diverse team that can bring added value not only through needed knowledge and skill sets, but also through their lived experience and perspectives.

While reading different definitions and explanations of new terms is helpful, having to apply and engage with various definitions concretizes learning. Therefore, the first teaching activity offers a structure for engaging in a discussion centered on the term *evaluation*. If you are new to the field, try applying the exercise with a few colleagues. When you feel more confident, facilitate the activity in a classroom, with clients or colleagues, or with a group of evaluators.

ACTIVITY 1.1. Learning through Discussion: Evaluation

Purpose: The facilitated discussion encourages people to think about how different evaluation definitions can influence a person's understanding, and therefore implementation and expectations, of evaluation.

Time: Allow approximately 5 minutes for preparation, 10 minutes for discussing the two questions, and 15 minutes for group discussion. Total approximate time: 30 minutes. Longer time frames are likely to encourage more in-depth discussions.

Preparation: First, select two evaluation definitions. Second, on flip chart paper, write the selected definition for each evaluation term, and put them on the wall (an option is to provide each group with both definitions). The groups need to see both definitions at the same time. Patton's and Scriven's definitions can be used, or other definitions as appropriate, such as the government department's, organization's, or donor's. Second, write the two facilitation questions, given in the next column, each on their own piece of flip chart paper, ensuring that only the first question can be seen. (Or write both questions on the same piece of paper, but cover the second question.) These two questions will be used during the group discussion.

Preparation of room: Divide the participants into small groups of three to five people, to allow for good discussions. Consider how the groups are organized. You may want to have random groups (e.g., all people with birthdays in January through March in one group, or all people wearing green), or more strategically thought-out groups (e.g., one manager, one evaluator, and one junior researcher). Consider what you want to happen in the groups. Do you want a discussion that mixes various perspectives in one group? Or a discussion among people who are likely to have similar understandings? Do you want one person to learn from another? Or do you want people to build relationships?

The exercise: Once groups are in their own spaces, ask people to silently read the two evaluation definitions. Then ask them to discuss the first question within their groups.

1. In what way are the two evaluation definitions the same? In what ways are they different?

Allow approximately 5 minutes for discussion, and then introduce the second question. It is OK if the group is still talking about the first one.

2. How might these differences influence an evaluative process or an evaluation, with regard to implementation and/or expectations?

Allow approximately 5 minutes for discussion. Then ask the members of one group to share their answers to the two questions, followed by the other groups. Facilitate a discussion that compares the responses. It is important to note that for some groups, facilitating the discussion immediately after the first question, before moving on to the second, may be appropriate; for other groups, that pace may be too slow. Some suggested probing questions during the larger-group discussion include the following:

Which definition do you prefer, and why?

What are the differences and the similarities?

Why is (or what makes) it important to have discussions that lead to shared definitions of evaluation, or at least an understanding that differences of opinion exist?

What might happen in an evaluative process if there was not a shared definition, or agreement to use the same definition, of the term?

Critical learning point: Different evaluation definitions will lend themselves to different expectations, which will likely create challenges during the evaluation. Discussing and developing a clear, shared understanding of evaluation and how the term will be applied are likely to result in shared expectations of the evaluation, which is an excellent place to start any evaluative journey.

Evaluative Thinking

Sometimes people use *evaluative thinking* and *evaluation* in the same breath. While they are indeed closely related, the terms connote different meanings. Evaluative thinking is what encourages the reflection on, and the use of, data and information. It encourages continual thinking about what can be done differently, improved, changed, or enhanced. I have heard some argue that anyone who can think evaluatively is an evaluator. For me, this argument is akin to contending that anyone who knows that it is important to brush their teeth twice a day is a dentist, or that anyone who knows that a steady diet of processed sugar is not good for the body is a nutritionist. Evaluative thinking is a critical part of being a thoughtful and engaged development worker or program manager. Thinking evaluatively means using data and information to be reflective, to learn, and to inform actions. While it is a critical element of being an evaluator as well, evaluative thinking does not by itself make someone an evaluator. (For a rich discussion of why this matters, skip ahead to Chapters 2 and 11.)

Monitoring

Now that we have discussed evaluation, and we have chatted about evaluative thinking, let's tackle the next term often associated with evaluation and evaluative thinking: *monitoring*. Monitoring is a common term used in public management that has crept slowly into evaluation books, journals, and evaluation conferences in the Western world. In the developing or emerging world, monitoring is often a familiar term.

We can all breathe a sigh of relief on how to define monitoring, as different groups tend to agree (though not always) on a common core definition. Monitoring is done on a day-to-day basis, and is the ongoing routine tracking of a program, project, intervention, or policy (or any of the other nouns mentioned earlier in this chapter). The confusion comes in when we are trying to distinguish monitoring from evaluation. Making this distinction is often a thorny conversation in the West, where monitoring can be subsumed into an evaluation process or completely disregarded as not being a part of the evaluation conversation. In developing or emerging markets, monitoring is so closely associated with evaluation that it is difficult to have a conversation that does not reference the M (monitoring) without the E (evaluation) in M&E. Speaking about monitoring as something different and separate from evaluation is practical, however, in that it allows a clearer understanding of what is done when, for what reason, and by whom. Let's take a closer look at monitoring.

Monitoring is something many people do naturally and answers a very basic question: **What is happening?** The question subsumes or includes three foci:

- What is (or is not) happening with the resources used for the intervention?
- What is (or is not) happening in terms of actions to implement the intervention?
- What is (or is not) happening as a result of those actions?

Sometimes we put in a number to gauge if something is happening: For example, did a person run 5 miles or lose 2 pounds? Setting these numbers, which evaluators have often called *targets,* helps us to specifically monitor progress, and show whether the person is moving toward or away from what he set out to accomplish. We discuss setting targets and other related concepts, such as benchmarks, more fully in Chapter 9.

Just a side note here: A critical element to remember in the evaluation journey is that staying on task is critical to untangling and focusing the process. However, because there are so many evaluation concepts and discussions that have interrelated and spin-off conversations, it is easy to get sidetracked. Sometimes a tangential term arises and causes havoc, and there is a need to stop and provide a detailed explanation. At other times, the tangential term adds a needed detail to the conversation (such as *target* above, or earlier in the chapter, when I stopped and explained the terms *data* and *information*) and requires no more than a quick, yet informative, chat. The critical part is, once the tangential term is explained, to return effectively to the core conversation. So let's go back to monitoring.

Let's assume we are monitoring an intervention. Our monitoring data suggest that the intervention is on track and moving toward where it intended to go, or maybe that it is not on track and thus requires one or more management decisions. The management decision could be to make an immediate change, do nothing and continue to monitor, or conduct an evaluation to find out why the intervention is (or is not) on track. The same monitoring data may also be used to demonstrate accountability. For instance, the data may be used in a monthly report that is presented to a board of directors.

Monitoring is a management function mainly used to determine whether an intervention is using appropriate resources and using them appropriately, doing what it aimed to do, and achieving the results it intended. Monitoring can also be used to keep an eye on external or internal factors that may influence the intervention (e.g., the economy, policy shifts, or people). Monitoring is often associated with informed decisions and accountability. Combined with evaluative thinking, monitoring provides a way to use data to learn and reflect; to improve a project, program, intervention, policy, or the like; to identify what was (or was not) achieved; and sometimes to recognize unexpected results.

BOX 1.7. Monitoring

When discussing the definition of monitoring, we need to consider four core concepts: The data collection is (1) empirical and (2) continuous, and data are used (3) to manage the intervention and (4) to provide accountability. Monitoring answers these questions: What is happening, and did the intervention do what we intended it to do? Monitoring does not ask if we are doing the right thing, or if we are doing things right; it also does not probe in any way to find out *why* the intervention did what it said it would do, or *why* it did not. Monitoring does not ask why or place a value on the results. When it starts to venture in that direction, it is more properly called evaluation.

Discussing monitoring is often more straightforward than discussing evaluation. However, monitoring also has its own nuanced definitions and various labels in different countries and in different organizations. For example, some people use the term *performance monitoring,* and others call the same process *performance management*. These terms have specific definitions that refer to tracking performance against stated targets, or goals. Earlier the term *M&E* was introduced, which can involve another fuzzy discussion, so let's talk about that next.

The Siamese Twins: M&E

My daughter, Gemma, has a book about a little girl who is told to count sheep to make her go to sleep (McQuinn, 2010). The little girl lies down and starts to count the sheep. Except the sheep do not behave and stay in a line: They play games, they swim, they hide, and they go in-line skating, all of which makes them very difficult to count. The little girl keeps running after the sheep to try to count them. Finally she exclaims, "You're exhausting!" and, ironically, falls asleep. I feel this way about how *M&E* is tossed around; it exhausts me. When someone says, "I am going to do M&E," I always wonder, "Well, which is it? Is it the M or the E?" Add to this that when I meet M&E experts, and they explain what they do, the explanations often vary considerably from one expert to the next. Oh, and then how the M&E persons define *monitoring* and how they define *evaluation* also often vary (refer to our earlier discussions on these terms). These are some of the conversations that exhaust me, and sometimes help put *me* to sleep.

When separated and defined, each term becomes useful, and then, when they are loosely linked back together, they serve a valuable role. However, when the term M&E is flung around without any questioning of what is being referred to, monitoring and evaluation become the troublesome twins indeed. Let's explore the practical link between the two terms.

The Practical Link

The link between the two terms is very practical. Monitoring an intervention should provide enough relevant data for a program manager (or whoever has the responsibility for the program, regardless of job title) to manage an intervention adequately and be accountable to those with power over the program (e.g., government, donors, community leaders). When something in that monitoring data raises a "why" question (e.g., "Why are people not coming to a training?" or "Why is that neighborhood having better results than other neighborhoods?"), evaluative thinking moves the manager toward evaluation territory, thus linking what is done through monitoring to evaluation. When an evaluator draws on monitoring data during an evaluation, it is important to verify that the data are available, accurate,

precise to the extent necessary, and credible. For an in-depth discussion on data and data credibility, please skip to Chapter 5.

BOX 1.8. Monitoring, Evaluative Thinking, and Evaluation

Monitoring data raises the flag; evaluative thinking makes us go "hmmm"; and evaluation answers why and brings the valuing.

Evaluation Triplets: Monitoring, Evaluation, and Evaluative Thinking

While the twins are troublesome, the triplets bring their own mischief. Here is a way to engage with the threesome. Monitoring is a process that answers the question about what is happening (descriptive). When we see that something is or is not happening, and we ask why, that moves us toward evaluative thinking. When we engage in further data collection or analysis processes to answer the why, and bring in a value judgment, that is evaluation. And often the processes associated with these words are all so tangled up into one exacting process that it may be difficult, if not impossible, to separate them easily in practice.

Now that we are near the end of our discussion on monitoring, evaluative thinking, and evaluation, I would be remiss not to introduce a few more acronyms that are becoming more common in the evaluation field. The newer acronyms I am seeing include MEL or MEAL (monitoring, evaluation, and learning), MER (monitoring, evaluation, and research), MERL (monitoring, evaluation, reflection, and learning), and PMEL (planning, monitoring, evaluation, and learning). And there are more. When working for or within an organization, we need to define and discuss their acronyms, what meaning each word brings, and who does what and when for whom, thus starting a conversation that leads to clarity and supports a useful evaluative process.

Alphabet Soup: The Use of Acronyms (or TUOA)

The evaluation field, and often the many technical fields it engages with (e.g., health, education, environment, community safety, climate change), seem to have a love affair with acronyms. In South Africa, nearly every day I hear or see the acronym ANC as an abbreviation for the African National Congress, the current ruling political party (often best known in other countries for being the political party of the late Nelson Mandela). When I started implementing health systems evaluations in South Africa, people talked about ANC all the time. I could not understand why the ANC played such a large role at the community health clinic level. Then one day I realized that ANC, in the health system, most often meant *antenatal clinic*—though sometimes it still meant the African National Congress. I was telling this story to a friend, and she said, "Oh, I thought it meant *antenuptial contract*." Hmm. My first piece of advice? Don't be shy; always ask what an acronym means.

My second piece of advice? Do not use acronyms if you can avoid them. First, it makes texts hard to follow, particularly for those not as familiar with the context as the evaluator or program staff. Second, in some cases, it can lessen the impact or take away the very meaning of what you are trying to convey. For instance, in a report about gender-based violence, using the acronym GBV can remove the very urgency or criticalness of the issue being addressed. It is a very different experience for the brain to read *GBV* rather than *gender-based violence*. A key point of writing an evaluation report (or anything, really) is to engage readers, not distance them.

While the term *research* is more thoroughly discussed in Chapter 2, the three words (monitoring, evaluation, and research) offer another set of triplets habitually intermingled in one conversation. Thus two activities are now provided to engage others in a discussion of these triplets and to provide foreshadowing for the more in-depth conversation about research and evaluation, and researchers and evaluators, in Chapter 2.

ACTIVITY 1.2. Monitoring, Evaluation, and Research Discussion

Purpose: The facilitated discussion brings out how the terms *monitoring, evaluation,* and *research* are often intermingled, and encourages people to think about how the three terms relate to each other in practice and how they are different. The activity works best as an introduction to the three words, not a summary.

Time: Allow approximately 5 minutes for preparation, 5 minutes for a discussion on each word in a small-group format, and 10–20 minutes for a large-group discussion. Total approximate time: 20–30 minutes.

Preparation: With a marker, write the word *monitoring* on a flip chart. On the next (and hidden) sheet of flip chart paper, write the word *evaluation*. Then write the word *research* (also on its own sheet of paper, and hidden). Thus, at the start of the exercise, the participants can only see the first word (monitoring). You may choose to write all three words on the same flip chart paper; if so, however, ensure that only the first word can be seen by the participants.

Preparation of materials: Each group needs three blank pieces of paper large enough to write a two- or three-sentence definition with marker or pen. It is often better to use marker, as then you have the option to attach the participants' work to the wall, and it is readable.

Preparation of room: Organize groups with no fewer than three members and no more than five. This size allows for good discussions. Consider how the people in the room are divided. You may want to have random groups (e.g., all people whose first names start with A–E in one group), or more strategically thought-out groups (e.g., one manager, one evaluator, and one junior researcher per group). See reasoning and hints from Activity 1.1.

The exercise: Before providing any definitions or explanations of monitoring, evaluation, or research, explain to the larger group that it is their job to teach a brand-new colleague (or classmate or staff member) the definition of monitoring. Show the group the word *monitoring* written on a piece of flip chart paper. Provide each group with one piece of paper, and have one person in the group write down the word to be discussed (i.e., monitoring) at the top of this piece of paper. Then ask them to discuss, in their group, how

they would explain monitoring to a colleague who is not at this training session. Ask the group to write down the explanation on the paper, below the word. When groups are finished, show the word *evaluation* to them, which is written on the next flip chart. Then have one person in the small group write this word at the top of a second blank piece of paper. Now that they have explained what monitoring is to the new person, ask them to explain what evaluation is to that person. Have groups discuss the term and write down how they would explain evaluation to the same colleague. When this is done, show the word *research* and repeat the steps. After each group has written a definition for each term, have the groups share how they would explain these three terms to their new colleague.

As a facilitator, listen for the distinctions made among the terms (or lack thereof). Ask the groups to provide feedback on their internal discussions, not just their definitions. After this exercise, present the more common definitions (those described earlier in this chapter, or other definitions relevant to the group) to close the session.

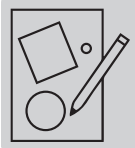
When facilitating, consider using these three questions:

- What were the challenges, if any, in describing these words?
- What is the importance of clarifying a distinction among these words? Why does it matter in practice?
- What are some of the likely challenges if the words are not clarified—for example, in common usage, in a job description, at the office, or in an evaluation process?

Facilitation hint: Before the groups present their descriptions of the three terms, tell them that they may only read what is on the cards. Reading what is on the cards helps the groups to focus, and helps keep the activity to its set time.

Optional exercise: To make this activity a bit more fun, use role playing. After all groups have their explanations of the three terms, invite one person to the front of the room to "play" the new colleague. Have the members of each group present their explanations to the person, and then ask the person which one was clearer. There is no intent for a winner; rather, the reasons for choosing one description over the others should invite more discussion.

Critical learning points: Knowing the definitions for these three terms, and practically applying them, are two different things. The group should be comfortable discussing the three different words and how they are practically applied in their organization or in their work (not just providing rote definitions).



ACTIVITY 1.3. Stories about Research, Monitoring, and Evaluation

Purpose: The facilitated discussion brings out how monitoring, evaluation, and research are often intermingled, and encourages people to think about how the three terms relate to each other in practice.

Preparation: None, unless you want to create your own story.

Time: Allow approximately 2 minutes to read the story, 5 minutes for individual thinking time, and 15 minutes for group discussion. Total approximate time: 22 minutes.

The exercise: Read the example. Here is an example that can be used: "When baking a cake, I first **research** what kind of cake I want to make—say, a chocolate layer cake or plain vanilla. I research differ-

ent recipes, and I choose one. I then set about making the cake, and I **monitor** what I put into it, such as how much flour or salt. I also **monitor** the cake to tell me how brown the top becomes, so that I know when to remove it from the oven. When **evaluating** my cake, I decide if the cake is good or bad, based on my preference, and sometimes I even listen to what others think of my cake." In my example, I have talked about using research, monitoring, and evaluation.

Now ask individuals to develop short examples (one per person) from their everyday life, and provide 5 minutes for the short examples to be written down. Ask for a few participants to share. These are some facilitation questions you can ask:

- What are the practical differences among the three terms?
- What challenges did you have in describing the three different processes, if any?
- How do these critical differences translate to your work?

Critical learning points: Again, knowing the definitions for these three terms, and practically applying them, are two different experiences. The group should be comfortable discussing the three different words and how they are practically applied in their organization or in their work (not just providing rote definitions).

Here are the more common conversations I have with regard to M&E. Reading these dialogues will build your confidence about engaging in these and similar discussions.

If someone holds the job of an M&E officer, or they say they are an M&E expert, what does it mean?

It can mean many things, so take the time to inquire what someone means when they use this term. While some organizations or governments may specify requirements associated with these job titles, there is no accepted or general job description for an M&E officer or an M&E expert. Someone holding this title or position (or self-labeling) may have extraordinary experience and knowledge of M&E, or may have little to no knowledge of either monitoring or evaluation.

Consider when someone takes a position as an M&E expert, and brings substantial knowledge and skills with monitoring to this position, but little to no evaluation knowledge or experience. This creates awkward situations for the M&E expert, who is then asked to do or manage work for which they have no background, training, or experience. At best, the M&E expert recognizes this challenge and seeks advice, training, knowledgeable others, or mentoring. At worst, the person's lack of knowledge and experience damages an intervention, gives a poor name to evaluators and evaluation, and frustrates key stakeholders and others who are negatively affected by a poorly implemented evaluation. So when people say they "do M&E," always ask them, "So tell me, what do you do, exactly?"

Finally, be aware that the same conversation may also happen when someone says they are an evaluator. Being an evaluator can mean many things—as described in Chapters 2 and 11 of this book.

I am not sure if I can ask an M&E person to do an evaluation. What do I do?

Clarify what skills, knowledge, and experience the person has with evaluation. Here are a few questions you can ask to determine the person's likely ability to do an evaluation that will meet your needs:

- What would be the steps that you would use to conduct the evaluation?
- What are your favorite evaluation approaches, theories, or models? Can you tell me about some you have used in the past?
- What are the ways in which you will determine the valuing criteria (i.e., determine their merit or worth), or what processes will be used to value the findings?
- Please tell me about the last evaluation that you conducted (or the most interesting, or the most challenging). What were some of the challenges? What worked well?

Answers to these questions will encourage a dialogue that provides insight into the M&E person's level of evaluation knowledge and experience.

Is a program manager expected to be an evaluator?

While the response to this question will depend on the organization, in general the answer is no. Program managers need "enough" evaluation knowledge to be able to engage with an evaluator, an evaluation unit (internal), or an evaluation team (external or internal), and/or to manage an evaluation. Here is another way to think about answering this question. I do not expect a program manager to be an auditor, or therefore to conduct a financial audit. However, I do expect a program manager to understand his budget, and to know if the intervention is over- or underspending. Similarly, I do expect a program manager to be able to use monitoring data. He needs to know enough to interpret the data and understand what the data are saying—for example, to take action (or not). For instance, in a high school tutoring program, the program manager should be able to answer questions such as these: Do children show up for tutoring (and, if so, how many)? Do children who show up for tutoring have better grades (if that is the intended result)? How many tutors are there who can cover advanced math, and does that number meet the current students' needs? Thus, just as a program manager is not (often) an evaluator, an evaluator is (in most circumstances) not an auditor or a program manager. A program manager needs to understand how to engage with monitoring and evaluation so that it is useful. Please see Chapters 2 and 11 for an in-depth look at the role of an evaluator.

Intervention, Activity, Project, Program: Doppelgangers or Differences?

I like the word *doppelganger*. It has a specific meaning that is exactly what I am trying to convey. Vocabulary.com (n.d.) provides a nice explanation: "Someone who looks spookily like you, but is not a twin, is a doppelganger. Originally, this was a type of ghost. The word doppelganger is German and literally means double walker—as in a ghost or shadow of yourself."

Doppelganger is a wonderful word to describe how the labels *intervention, activity, project,* and *program* are used. For example, *intervention* and *program* are sometimes used interchangeably to discuss the larger picture (e.g., "We have an environmental program," or "We have an environmental intervention"), while *project* is sometimes but not always used to describe a collection of smaller activities (e.g., "Our program focuses on projects that address fish and wildlife conservation"), and *activities* are often specific (e.g., "Our project has several activities, one of which is to issue licenses to fish farmers"). And sometimes not. When you are working with any organization, ask them to describe how they use these labels and terms, so that everyone is clear about what is what, and which label is attached to what description.

Here is an example of how these words may be used in Organization A.

Intervention	Activity	Project	Program
The nonprofit is teaching high school students how to study for exams.	One tutor works with one student for 2 hours per week.	Each site (and there are 450 sites) has five tutors who work 5 days a week for a total of 20 hours.	It is a national program for inner-city youth.

Here is an example of how these words may be used in Organization B.

Project	Intervention	Activity	Program intervention
The nonprofit is teaching high school students how to study for exams.	One tutor works with one student for 2 hours per week.	Each site (and there are 450 sites) has five tutors who work 5 days a week for a total of 20 hours.	It is a national program intervention for inner-city youth.

Vision, Mission, Goal, Objective: Management Speak Translated into M&E Speak

For *mission, vision, goal,* and *objective, doppelgangers* is an appropriate term to reuse here, and let me just say "ditto" to what I have written in the previous section. The management language and the M&E language can be thought of as distant cousins (perhaps second cousins),

or we can view them as terms that live in their own separate yet parallel universes. When you are working with management terms, identify what should happen or be achieved when, and to what extent, the organization is held directly accountable for which achievement (i.e., which result). Knowing both (what is intended to be achieved and what the organization is held directly accountable for) will clarify what *needs* to be monitored and evaluated (there are many things that can be monitored, and we delve into this in Chapters 3 and 4), and what is the far-reaching but necessary dream (e.g., a better world, equality for all), which is not (often) assessed in the lifetime of the intervention. I say "necessary," because most interventions are aiming for something just beyond their reach, and this dream guides the work in the present, like a guiding star. When M&E words are mixed and matched with management words, they tend to (though they do not always) have corresponding levels. It is with trepidation that I note the following common equivalents:

- Vision, mission, goal = Impact
- Objective = Outcomes

Again, these are often somewhat similar levels—but not always. Please jump ahead to Chapter 7 for a more in-depth look at why these examples are provided with some trepidation.

Discussing Concepts, Terms, and Labels

Do not assume that everyone engaging in an evaluation discussion has the same understandings of concepts, terms, and labels—even those that have concrete definitions. Here are three pieces of advice:

- Discuss each concept or term.
- Define and describe how it will be applied.
- Write it down and get physical sign-off on it.

WRAPPING UP

Common evaluation concepts and terms, even when they have concrete definitions, can be defined, interpreted, and applied by different people, groups, and organizations with nuanced or astronomical differences. Knowing the difference between concepts and terms that have specific definitions and uses, and those that do not, is important. An evaluator needs to know when she can say, "That is indeed the definition; let's discuss how we will use it," and when she needs to negotiate a concept's or term's definition, meaning, and application. When an evaluator ensures that those who need to be on the same page with regard to a term or concept's meaning are actually in agreement, she avoids mayhem, misunder-

standings, frustration, and likely major disappointment in the evaluation and its process. An evaluator who explicitly engages with evaluation words, concepts, and their meanings (whether through a 2-minute conversation, a written clarification, or a longer process) is more likely to have a useful evaluation. Some side benefits of clarifying language (evaluators call these "unintended results") may include building capacity, strengthening relationships among those engaged in the process, and leveling the playing field among stakeholders.

The next chapter moves us from talking about monitoring, evaluative thinking, and evaluation to another topic that is often as intensely discussed: the differences between evaluation and research, and between an evaluator and a researcher. Before you leave this chapter, however, join me in a conversation.

Our Conversation: Between You and Me

Here are two types of situations that you may encounter with regard to the concepts and terms just covered. Practicing with these common scenarios can be a great help as you deal with these kinds of situations in the real world.

1. Your clients (or boss or colleagues) have asked you to do the M&E for a program. You have no clear idea of what they want you to do. Start at the beginning, as we have done in this chapter. Encourage a dialogue to get clarity about what they expect from M&E. To prepare, identify the organization's or client's definition of *M&E* (or its component terms, *monitoring* and *evaluation*), or any description of how the acronym is used in the organization. Reading the organization's information will help you identify a basic place to start asking questions. Here are some general discussion guidelines:

- Be clear about what is within your boundaries and what is outside them, with regard to what you are capable of doing in terms of your own knowledge/skills and the resources that are provided.
- Move the initial conversation away from labels. For example, if the request is "We expect you to do monitoring," a good follow-up question would be "Could you explain a bit more what you mean by *monitoring,* perhaps with some examples?" If they seem hesitant or unsure, offer some suggestions or examples of types of activities often associated with monitoring.
- Seek to understand how the data that you collect, and the processes that you engage in, will be used by whom to do what.
- Once you think you have a clear understanding of what is wanted, provide some concrete descriptive examples of what you are going to do, and have them react to those examples. Provide some examples as well of what you will not do, and have them then react to those examples.

Draw on the definitions, discussions, and exercises demonstrated in this chapter to engage your clients (or boss or colleagues).

2. You are working in a place where English is not the first language, and there is confusion about how to translate or how to use the terms *monitoring, evaluative thinking,* and *evaluation*. A discussion of the distinctions among *monitoring, evaluative thinking,* and *evaluation* is not always a useful conversation when terms are translated. Or the terms may simply not be useful in the culture in which you work, even if English is the dominant language. So remove the labels. Facilitate a process that draws from how monitoring and evaluation are described in this chapter, without using the terms. For instance, ask the person, "How will you know if the intervention is doing what it is supposed to do?" What *is* needed is a clear understanding of what data are collected when, by whom, how, and for what reason, and how those findings will be valued and by whom. Once all that is clear, you can then discuss who does what in that process. Then and only then, choose appropriate labels (e.g., "We will call this *X* and that *Y*"), *if* labels are needed. For example, labels may be needed for the group to engage with their donor or others outside the organization. Or perhaps no labels are needed.

The other day my neighbor said to me, "Tell me again what you do for a living. You are a researcher, right?" Sigh. Please come join me in Chapter 2.

CHAPTER 2

The Tale of the Researcher and the Evaluator

I am a researcher; therefore, I am an evaluator.

The conversation that informs the evaluator–researcher debate is rooted in the "research versus evaluation" debate, so let's start there. Sometimes it is a deliberate and enjoyable conversation, and sometimes it is a subtler, strained one. Either way, the extent to which differences between research and evaluation exist influences several key decisions, such as what to study to be an evaluator, or whom to hire to design, implement, or critique an evaluation.

Before we embark on a somewhat dizzying journey, keep in mind that the basic aim of social science research is to produce or generate knowledge. Patton (2008) says that evaluation aims to influence actions; it is action-oriented. He notes that some evaluation texts mingle these definitions, such as Rossi and colleagues' (2004) popular text *Evaluation: A Systematic Approach,* which is in its seventh edition and is used in many social science graduate programs. In their book, Rossi and colleagues define **program evaluation** as "the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions" (p. 16).

Some people put the two terms together and talk about **evaluation research.** Kelly (2004) describes evaluation research as an approach in which the purpose is to "solve practical problems" (p. 522), suggesting that the focus is not on judgment or improvement of interventions. Babbie (2017) notes that the purpose of evaluation research is to "determine whether social interventions or programs have had their desired effects" (p. 387). Here, the word "value" is noticeably absent. At the same time, other evaluators use the same phrase, *evaluation research,* to mean research that is done *on* evaluation. And let's not forget basic

and applied research. McBride (2013) tells us that basic research aims to address "fundamental processes of behavior," while applied research aims to "solve real-world problems" (p. 10). Hmm. Obviously, distinguishing between an evaluator and a researcher can be a bit mystifying—so I will make it as clear as I can. Keeping Patton's understanding of evaluation (its aim is to influence actions) and the core meaning of social science research (its aim is knowledge generation) in your mind, let's take a closer look at the differences between researchers and evaluators.

Why It Is Important to Understand the Differences between Researchers and Evaluators

Understanding the difference between evaluators and researchers can inform your career path in terms of which one you may like to be and therefore what to study. It can also clarify for you which types of positions to apply for, or whom to hire. While I was writing this book, an international development agency advertised for an evaluator to review evaluations of development programs in education and health from an external perspective, and to provide critical feedback. The job description asked for a person with strong communication skills, strong qualitative and quantitative research skills, and knowledge of the organization—all of which gave me pause. What about knowledge of evaluation methods, theories, or approaches? Or practical experience in designing or implementing evaluations? Or knowledge of how power, politics, culture, and language influence an evaluation approach? I was a bit aghast; surely they did not want a researcher to review an evaluation and provide feedback? Or maybe, just maybe, they did not know the difference between a researcher and an evaluator?

Researchers and Evaluators: Compatriots or Competitors?

Having a conversation about the differences between evaluators and researchers can be enlightening and uncomfortable at the same time. Sometimes the conversation is sensitive, such as when a researcher considers himself to be an evaluator. Or, awkwardly, you are a researcher and are wondering, "What does an evaluator have that I do not?" At other times, it is frustrating, such as when an evaluation commissioner hires a researcher to design an evaluation. So should you ever find yourself engaged in these kinds of ponderings, conversations, or situations, either with yourself or with someone else, consider these three essential conversation points: **valuing, purpose,** and **approaches.**

• *Valuing.* Valuing findings is at the core of an evaluator's work. Evaluators determine criteria (which can be done in multiple ways; see Chapter 12) to value evaluation findings. Researchers are not asked to value their research findings.

Wait, stop on that first point. I find some research to be very valuable.

I agree that some research is very valuable. Researchers use data to support a claim about something; however, they do not follow up that claim by valuing it. Here's an example. All females on my mother's side of my family suffer from migraines. Research has shown that certain so-called "trigger foods" may bring on migraines, and research also shows that certain types of drugs alleviate the pain. The females in my family are delighted about the research findings. While the researchers did not value their findings, my female family members found the research to be very valuable. For the evaluator, valuing a process or result is at the heart of her work.

• *Purpose.* Researchers ask questions on behalf of the larger scientific community. Basic scientific research seeks to uncover new knowledge, investigate and test theories, identify findings, and generalize them. Evaluators aim to provide information to a specific group on a particular intervention. The information is intended to be used to improve or judge that intervention—to learn more about that particular intervention, in a particular context, for a particular group of people, in a particular time frame.

• *Approaches.* Evaluation uses research methods, and then also brings its own approaches, theories, and models that are specific to the evaluation journey (see Chapter 15). Because evaluation approaches, theories, and models guide the evaluation process and its valuing framework, an evaluator needs to know them; a researcher does not.

BOX 2.1. Research and Evaluation: Some Distinct Differences

While the aim of applied research (and researchers) is often to generalize the results to the larger population (McBride, 2013), the aim of an evaluation (and therefore evaluators) is more commonly to look at a specific intervention, program, process, or policy and to judge its merit, worth, and significance, with the intent that the findings will be used by a specific group to make management decisions. A researcher fills the role of methodologist; an evaluator can often fill the role of methodologist, evaluation theorist, facilitator, negotiator, educator, and more. See Chapter 11 for further elaboration on an evaluator's potential roles.

Let's take a moment to look at what a conversation between a researcher and an evaluator might look like.

I am a researcher.	I am an evaluator.
I seek to generate knowledge, and I answer questions. I contribute to broader scientific knowledge.	I seek to generate knowledge, and I contribute knowledge to use in making decisions or improving an intervention or policy, for a specific group of stakeholders.
I use research methods of inquiry.	I use research methods of inquiry. I also use evaluation frameworks, theories, and approaches.
I identify results.	I identify results, or sometimes I use a researcher's results. And then I often work with diverse stakeholders to determine criteria on which to value those results.
My work is researcher-centric. I get to decide what I research and how to focus the research, though sometimes others, like those who fund me, also influence the decision.	My work is stakeholder-focused. Sometimes I help to shape stakeholders' questions. I shape them so that answers are likely to be useful to the stakeholders.
I publish my results in journals and books, and I also share my research at conferences so that the knowledge generated can be used.	I provide my reports and findings to the key stakeholders, so the findings can be used by them. Sometimes I publish, although publishing is not key to my evaluation work. Sometimes I give presentations at conferences on the work that I do.

Both Chapter 1 and this chapter may appear to suggest that conversations to clarify words and how they are used (e.g., **evidence** or **data, evaluator** or **researcher**) are easy conversations. They can be. At other times, however, to say that they become more antagonistic is to put it nicely. Knowing how to facilitate these tense conversations is a skill essential to being an effective evaluator, as difficult conversations can arise in all parts of an evaluative process. Let's stop for a moment and talk about facilitating useful conversations.

How to Facilitate Useful Conversations

In most evaluative processes, there will be times when people are on opposing sides, such as when I say "researcher" and you say "evaluator." Practicing how to manage these discussions in often less sticky situations (e.g., evaluation vs. research) can prepare you for occasions when you will need to engage with facilitating more challenging ones. Here are some other likely examples: (1) In an evaluation, there are data to answer a question that are as convincing for one side as another (e.g., "Yes, do it" vs. "No, do not do it"); (2) two groups (or more) analyze the same data and reach different conclusions; (3) there is agreement on the

conclusions, but a difference of opinions on the recommendations; and (4) there are people who just plainly have different opinions or understand a situation or word differently, and to whom the group needs to listen.

The approach described next describes how to facilitate a process that encourages people to listen to opposing viewpoints. There are three reasons to consider using this approach. First, when people know that others will listen to them, they are more likely to engage and work with these others, even when they have different opinions or assumptions. A second, and somewhat related, reason is that it can be very frustrating to perceive (rightly or wrongly) that no one is listening. This can cause a person or group to "shut down," and thus can cause an evaluator to miss important perspectives in the discussion. Third, truly listening to what others have to say who think differently opens new ways of understanding.

ACTIVITY 2.1. Listen–Speak–Listen

Purpose: A useful approach to engage groups who have very strong and clearly very different viewpoints is to facilitate a discussion where people *must* listen to other ways of thinking and other ideas. The purpose is to ensure that each idea, thought, or fact is spoken (or signed, as in sign language) and that people listen to each other; the purpose is not (necessarily) to reach agreement.

Time: The time will vary considerably, depending on the number of people in the room, and the number of ideas, thoughts, and/or facts that people want shared. I recommend, however, that groups be allotted 10–15 minutes to write down their key ideas or thoughts. While topics and the size of the group will further determine the time frame, keep in mind that it is hard work to listen. Limit the discussion, if possible, to 15–20 minutes. A longer-drawn-out listening session can be counterproductive, as people tend to stop listening when they are tired. Total time is approximately 25–35 minutes, though it can vary.

Preparing: Prior discussions inform the debate topic. Prepare small pieces of paper (with space to write one or two sentences) for each group; there should be at least one piece of paper per person, though more pieces of paper are recommended (should there be more ideas). Gather pens or pencils.

The exercise:

1. Summarize the question or debate on a piece of flip chart paper, and place this sheet at the front of the room. Summarize the two (or more) sides as equally as possible. Read these to the group.
2. Have people choose the side of the debate that they support more (or most) strongly, and form groups. If any persons are truly neutral or undecided, or just uncomfortable with the exercise, they stand off to the side as observers. During the activity, they are not allowed to speak or sign.
3. Read the rules to the group (see "The rules" on page 34).
4. Provide each group with pieces of paper and some pens or pencils, and ask them to write *one* key idea/point per piece of paper. Each person in the group should have *at least* one slip of paper to present, if possible, though each person may have several pieces (i.e., several ideas).
5. When the groups are ready, have them form a circle, with an opposing group member on either side of each person, so that every other speaker is on the same team. To decide who goes first, flip

a coin and choose a person from that group. He reads his statement, and then the person to his left reads hers, and so on, until all the points are made. Sitting in a circle encourages a discussion-like format. The process continues until all people are heard.

The rules: There are six simple rules.

1. People must show respect at all times (i.e., no yelling; no disrespectful words or body language).
2. Only one person speaks at a time.
3. Each person may only read what is on the paper—no more, no less.
4. Once an idea, thought, or fact is shared, it cannot be repeated, even if different words are used.
5. There is to be *no response* when the idea is presented, from either the other team or the observers.
6. Everyone must agree to the rules before the discussion starts.

This exercise encourages listening that enables engagement with other points of view. Such engagement has the potential to inform deeper thinking, promote stronger conclusions/more focused recommendations, or lead to agreement—although, of course, it may not. The Listen–Speak–Listen discussion is an example of a situation where an evaluator fills the role of negotiator and facilitator. (Again, for more on an evaluator's roles, see Chapter 11.)

Here is a specific example of when to use the Listen–Speak–Listen activity. Empirical data sometimes provides two opposing answers to the same evaluation question. While the evaluator can present both empirical findings to the evaluation user (and sometimes it is acceptable to do so), it can sometimes leave the evaluation user confused at best, or annoyed at worst. After all, an evaluator was brought in to provide information to inform management decisions, not a response much akin to "Well, sometimes it does, and sometimes it does not." While both sides can still be presented in the report, reasons (drawn from the Listen–Speak–Listen discussion) can now be given for why one answer is likely to be more appropriate than the other (through the depth of additional detail and explanation), or at the very least the evaluator can provide more information that management can use to make a decision.

Evaluation versus Research

- ***Have a minute?*** If you want a good giggle, peek at a blog by Patricia J. Rogers and E. Jane Davidson called Genuine Evaluation (*www.genuineevaluation.com*). For example, Rogers and Davidson provide a funny, realistic take on the difference between evaluators and researchers with their May 25, 2012, list of "Top ten things you'll never hear from the researcher you hired to do an evaluation."
- ***Have a few hours?*** Read an article by Levin-Rozalis (2003), "Evaluation and Research: Differences and Similarities," which can be found online in the *Canadian Journal of*

Program Evaluation. Or read a chapter by Mathison (2008), "What Is the Difference between Evaluation and Research—and Why Do We Care?" in N. L. Smith & P. R. Brandon (Eds.), *Fundamental Issues in Evaluation,* pp. 183–196.

- ***Have some more time?*** Read Chapter 2 in Michael Quinn Patton's (2008) *Utilization-Focused Evaluation,* fourth edition.

Researcher or evaluator? Research or evaluation? As we have reviewed in the chapter to this point, there is a difference. Understanding this difference should influence many things—from what knowledge and skills to pursue, to who is hired, to how something is shaped and used. The differences are confounded when researchers self-identify as evaluators, or when evaluation is mislabeled as research (and vice versa). And more mixed signals are sent when, for an example, an economist does an evaluation, yet labels herself as an economist rather than an evaluator. Knowing and understanding the differences (and the points where research overlaps into evaluation) is critical for evaluators, and those who commission them, to ensure that persons with the right knowledge and skills are present to conduct an evaluation.

The evaluator-versus-researcher conversation has opened up a space to confess that in the real world, some conversations may not go smoothly. It has also suggested a practical solution: the Listen–Speak–Listen activity, which provides a technique for facilitating lively conversations and active listening. This approach can be used at any point in the evaluation process when, for example, different opinions arise, where data do not provide clear answers, or when different interpretations of the data offer useful but opposing insights.

Now that we have covered some basic differences between evaluation and research (and evaluators and researchers), and have learned a practical facilitation technique, it is time to commence our evaluative journey. Before we let the adventure continue, however, I invite you to have a conversation with me.

Our Conversation: Between You and Me

Here is a situation that you may encounter with regard to the concepts and terms we have just covered. Practicing with this common scenario can be a great help as you deal with this kind of situation in the real world.

If someone asked you the following, how would you answer: "Are you a researcher or an evaluator?" Think about how you might respond so that a nonresearcher or non-evaluator would understand you—that is, without jargon, technical terms, or abbreviations. For example, most people would probably start drifting off if you were to say, "Well, I am an evaluator, and I do empirical research that then draws on evaluation theory to value findings—and a researcher, well, a researcher uses social science theory to . . ." If we met at a cocktail party, and I asked you that question, what would you say to me?

Describing the Evaluative Journey

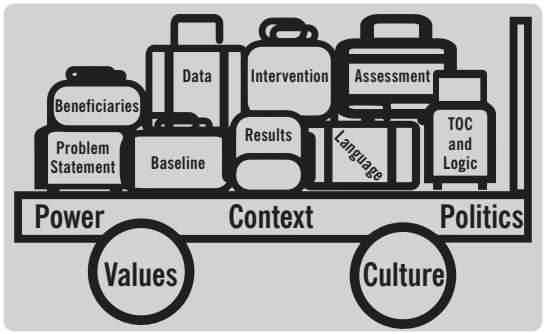
The next eight chapters illustrate the iterative logic of monitoring and evaluation, and suggest a process that is appropriate for nearly any monitoring or evaluation exercise. There are challenges to writing a book about an *iterative* process (i.e., one involving repetition to achieve a series of outcomes). First, it is nearly impossible to write iteratively without sounding redundant or having too much repetition for a reader. Although an iterative process works well in a workshop or discussion, and in real-life evaluative processes, it does not work so well in a book. The second challenge is that sometimes what logically seems like the beginning, or the place to start, is not. Sometimes it is better to dive into the middle of the process, in order to untangle the beginning and sort out the end, which brings us back to the beginning and then back to the end, and then back to the middle. And so on. Related to this is that your evaluation journey may start in a different place than where the book starts the journey. The third challenge is that there are often concepts, ideas, and definitions buried in each step that need to be discussed. So you see my dilemma: how to present the next several chapters—which by nature are iterative, with different potential starting places that can spiral off in multiple directions (depending on the evaluator and the context)—so that the steps make sense to you, my readers.

The Suitcases for Our Journey

The solution is to organize our journey into a trip where I bring along suitcases that keep the process organized by themes. The suitcases represent the themes that will be covered in each chapter; thus you will know what I am covering, in a broad sense. We will carry our suitcases with us, unpack what is needed, and then tuck them away until, once again, we need something that is inside. For instance, the suitcase labeled "Assessment" is packed to the brim with ideas, discussions, and definitions of such terms as *indicator, facts,* and *methods*. However, I will only take out of the suitcase what is needed at that point in the evaluative process; thus the same suitcase will show up multiple times, and each time something else may be pulled out, or quite possibly the same concept or term will be reframed as needed for that part of the journey. The journey can be used to:

- Reconstruct what happened (backward-looking).
- Construct what we hope will happen (forward-looking).
- Support implementation or innovation as it happens (real time).
- Inform the completion of a practical M&E framework.

The journey is thus used to provide answers to, and inform decisions about, what and how to assess and value an intervention—in other words, how to monitor *and* evaluate. All the topics covered during our evaluative journey are depicted on the next page.



The Journey's Illustrative Example

Throughout the remainder of this book, an example of a girls' education intervention is used repeatedly, to provide a consistent demonstration of how to apply concepts, terms, and processes. The girls' education intervention will unfold along with the journey, and will continue to be refined in each step, as an intervention would in an actual evaluative process.

POTENTIAL USERS OF THE EVALUATIVE JOURNEY

The rest of this book describes a journey through an evaluative process that is mindful of how power, politics, language, culture, and values shape that journey. The process has many iterative, interconnected, and interreliant steps. These include describing and understanding the intervention, defining and interrogating the reason for the intervention, determining what to assess, and gathering data (the monitoring). The information is then used to reflect and learn and to inform an evaluation design and process that increases broader knowledge, aids in judgment, supports learning, and/or otherwise informs decisions. The evaluation journey is designed to have multiple uses for six specific groups:

- *Students of evaluation and emerging evaluators.* Students of evaluation and emerging evaluators can use the thinking and process in the remainder of this book to improve their practical understanding of evaluative processes. Immersing themselves in the book will enable them to understand how and where evaluative thinking starts (and that there are various starting places), how the intervention's logic and theory are intertwined, and how that relates to and informs evaluation. The book grounds students' and emerging evaluators' understanding in how to engage with and construct effective monitoring frameworks, and ultimately how to design and implement useful evaluations.

- *Implementers of interventions.* Implementers are encouraged to make informed decisions that support achievement of an intervention's goals. The thinking and its processes

can be used so that implementers and their colleagues have a clear, holistic description of an evaluable intervention, and can use that understanding to engage in the M&E process in a practical manner. Furthermore, the process supports implementers as they engage with evaluators, donors, and others to whom they are accountable or with whom they otherwise interact.

- *Managers of interventions.* Managers of interventions can gain clarity on the intervention and the evaluative process. This clarity can support needed insight into how and why stakeholders with different perspectives may have different expectations regarding what is assessed, how it is assessed, and how findings are valued. This understanding can then be used to encourage constructive and mutually beneficial dialogues with implementers, donors, evaluators, and others to whom the managers are accountable or with whom they otherwise engage.

- *Evaluation teams.* Evaluation teams can use the process to clarify an intervention, which then informs refinement of evaluation questions, the evaluation design, and its valuing framework, and helps determine how and to whom to communicate the findings. Importantly, the process encourages teamwork with a shared understanding on all aspects of an evaluation process.

- *Lone evaluators.* A lone evaluator can use the process to think. When there is no one to engage with, and the evaluator needs to think about how to engage in an evaluation, going through the process will help keep her thinking organized and her decision making explicit—both of which are needed in an iterative evaluation process.

- *Confused evaluators.* The confused evaluator is provided with a way to untangle, engage with, and sort out confusion. Nearly all of us have all been there—perplexed, flummoxed, or mystified by the evaluation process. It happens. For example, someone says, "I need you to look at second-level outcomes and also assess the broader impact, and consider the immediate outcomes, and also look at efficiency, effectiveness, impact, and sustainability. And gender. For the capacity-building program." Our immediate response may be, "Say what?"

WRAPPING UP

We have begun the evaluation journey in Chapter 1 and this chapter by clarifying the critical use of language in evaluation, and distinguishing between an evaluator and a researcher (and research and evaluation). Chapters 3–10 present an iterative process that can have different uses for different people. A girls' education intervention, the suitcase symbols, and the clarifying icons will guide the iterative evaluation journey for all its different users. Grab your hat and shoes; we're off on our evaluation adventure!

CHAPTER 3
Starting the Evaluative Journey

We often forget to draw a new picture because we are so busy criticizing other paintings.

—DEBASISH MRIDHA, American physician, philosopher, poet, seer, and author

Now that we have clarified some basic evaluation terms, and have taken our suitcases in hand, let's talk about one place to start an evaluative process. And no worries: A conversation about monitoring will naturally be wrapped into the discussion. Whenever we are serving as evaluators or working with an evaluative process, it is critical for us to thoroughly grasp and explain the intervention with all its complexities, obscured and otherwise. While each chapter delves deeper into, and continues to untangle, the intervention and the evaluative process, Chapter 3 provides one place to start: the reason(s) for the intervention's existence.

I am going to share a trade secret with you: Most evaluators do not instinctively understand an intervention in its entirety, or in all its intricacies and nuances. Even if an evaluator is an expert in a technical field within evaluation (e.g., a specialist in education evaluation), she rarely (if ever) knows automatically, and immediately, exactly what or how to evaluate an intervention. Evaluators are not omniscient (and be very wary of ones who claim to be!). Evaluators need to have a clear understanding of *exactly* what an organization is doing or planning to do (the **intervention**); to fix, change, or make "it" better ("it" being the **problem statement**); and for whom, exactly, the intervention is intended (the **beneficiaries**). And that is just the *start* of what evaluators need to know. Now that this trade secret has been shared, let's delve into these three concepts, starting with the intervention.

Introducing the Intervention

Despite what seems to be a common belief, an evaluator cannot magically answer the question "How do we evaluate it?" or "What did it accomplish?" without first knowing what "it" is that is being monitored or evaluated. The clearer "it" is, the easier it is for the evaluator to answer these and other questions.

BOX 3.1. Using the Term *Evaluative Process*

The term **evaluative process** is used to describe the entire flow from inception of the intervention, to what happens, to the valuing of what happens or does not, to the sharing and use of that information, and everything in between. The process and discussions described in the book can occur before an intervention starts (forward-looking, the planning phase); as the intervention unfolds (in real time, or for an innovative or developmental process); or after the intervention or components of the intervention have taken place or ended (retrospective).

What Is the Intervention?

Whether designing an intervention so that it is evaluable, monitoring an intervention as it is being implemented, or conducting an evaluation of the intervention, an evaluator first needs to understand what the intervention *is.* (See Chapter 1, pages 11–12, for a full listing of other nouns that can be substituted for *intervention*.) There are a few exceptions, such as in developmental evaluation, where the purpose is to inform and document the intervention as it unfolds; see Chapter 15 for specifics. The basic details that are often needed to understand an intervention, and therefore the basic questions that are usually asked, focus on (1) what it does, (2) what it aims to fix, change, or otherwise improve, (3) what it aims to accomplish, (4) how it intends to accomplish that, (5) where, and (6) for whom. Additional useful information could include (7) who provides the funding, (8) how long the intervention has been in place and for how long is it likely to be in place, and (9) who implements the intervention. These questions guide the early search for information about the intervention.

Ways to Obtain Data about the Intervention

Having conversations, reading documents, and observing (i.e., going to see something), are initial ways to obtain data describing the intervention. While the following pages provide a *brief* introduction to the types of data collection often used to learn about the intervention, Chapter 5 offers a more detailed discussion on data collection.

Reading Documents (Conducting a Document Review)

One way to obtain data describing the intervention is to read documents. The purpose of this process, also called a *document review* or *archival document analysis,* is to help inform an understanding of the intervention or answer a question in the evaluative process. The documents can contain words, numbers, or pictures (e.g., videos, drawings, or photos). Documents can be sourced from those who fund the intervention, obtained from implementation staff, requested from managers, or found on the internet, to name a few sources. Types of documents may include (to name just a few examples) pamphlets, funding proposals, meeting notes, business plans, quarterly or monthly reports, newspaper articles, promotional videos, theories of change (see Chapter 8 for a description of such theories), monitoring frameworks, or previous evaluations. An organization or intervention's website or webpage is often a very useful place to get started, as it likely provides a short description that gives a sense of the intervention, as well as other documents. However, the website or webpage should not be the sole source of data or information. Although it may provide easily accessible information or data, these may also be outdated or may only provide a partial picture.

Having Conversations (Interviews)

One way to understand what the intervention is, and why it exists, is to ask different people who probably have different perspectives on and insights about the intervention. As a starting place, consider talking to people who implement the intervention (perhaps those who designed it) and those who manage it, either individually or as a group. You can also talk to those who benefit, receive services, or otherwise take part in the intervention. These conversations are called *interviews*. Interviews can use open- or closed-ended questions.

An open-ended interview question is one that has no preselected answers. An example of such a question is "Tell me about your experience of reading this book." The upside? A person can say anything, thus providing a wealth of information. The downside? A person can say anything, thus providing broad and perhaps haphazard (and not useful) information. A more focused yet still open-ended question can also be used, such as "Describe what you like about how M&E is discussed in the book." That question guides the person to talk about what they liked about how M&E was discussed, not just any topic in the book. There are advantages and disadvantages of using more focused (sometimes known as leading) questions. The advantage is that it increases the chance of obtaining usable answers for description. The disadvantage is that respondents may be more likely (either consciously or unconsciously) to provide an answer they think you want to hear, and perhaps not mention another topic that would have proven informative.

An interview can also use closed-ended questions, which may or may not have predetermined answers. An example of a closed-ended question is "Do you consider yourself an evaluator?" which can be answered "Yes," "No," or "Not yet." Advantages of closed-ended

questions are that they are often easier and quicker to ask and answer. A potential glitch is that someone who brings no knowledge can still answer the question, even if the person asking the question does not realize that; similarly, a person who does not understand the question can still likely provide a response. Furthermore, if there are preselected answers, the *actual* answer a person wants to give may not be a choice.

The evaluator can mix question types as needed. Finally, an evaluator can ask follow-up, open-ended questions to either an open- or closed-ended question, which are called *probes*. An example would be "Tell me, what makes you think that?"

Going to See the Intervention (Observation)

Observing the intervention can be a good way to begin to understand what the intervention does. Before an observation can take place, however, an evaluator will need to ask about and adhere to any applicable guidelines, such as obtaining ethical or legal permission, understanding what is culturally appropriate, and considering practical things such as determining appropriate times for visits. Observation can be done with a checklist that will provide focus (e.g., a set of closed-ended questions as described in the previous section); a partially structured guide that gives some guidance on what to observe, such as thematic areas (e.g., male–female interaction, use of cellphones); or an exploratory approach that provides no guidance (the evaluator uses just a blank journal or the electronic equivalent to take notes). Typically, an evaluator will observe how participants and staff interact, what the intervention does (e.g., tutoring, training), or how participants gain access to the intervention (e.g., is it near a bus stop or in a central location?), though observations can be broader. The chief advantage of observation is that it immerses the evaluator into the field, providing a sense of the intervention (e.g., its context and meaning) that often cannot be derived from interviews or document reviews. A disadvantage, however, is that when people know that an evaluator is coming, a different picture may be presented from what happens normally. In addition, observation can be costly and time-consuming.

Gathering and Organizing Data for a Nutshell Description

In interviews, when people respond, their words (their own or preselected ones) are called **data.** In the document review, the words and pictures gathered are also called *data*. In observation, what is observed is also called *data*. (For a reminder of the distinction between *data* and *information,* see Chapter 1, page 13.)

Here is one example of the steps to consider when you are starting an evaluation and you need to learn about an organization or an intervention. Visit the webpage or website (if one does not exist, move to the next step). See what kinds of program descriptions and documents exist. Then call the organization and ask for additional documents not currently on the website that would be useful in understanding the intervention. Give some

examples from the list provided on page 41 (e.g., newspaper articles, quarterly reports, evaluation reports). Read the documents, and sketch a description of what is known thus far. Then write down any areas that need clarification, and use these to inform the interview questions. Share the questions with the organization (by phone or email); then ask who is best situated to answer these questions, and organize the interviews. The interviewees (also called *respondents*) may or may not be able to answer the questions. If not, that is fine; continue to ask the unanswered questions throughout the evaluative process. Your initial sketch of the intervention is what is called the *nutshell description* (defined in the next section). This is just one approach of many that can be used to grasp the initial understanding of what "it" is that we are seeking to monitor or evaluate.

The Nutshell Description

Synthesizing data from interviews, document reviews, and/or observation provides a broad initial understanding of the intervention, called a *nutshell description*. Such a description is needed to begin to understand what the intervention is aiming to accomplish, how it intends to accomplish that, where, when, and for whom. It is called a *nutshell description* because of its brevity; ideally, the description would fit into a proverbial peanut shell. In the early stages of the evaluative journey, it is not likely that all the intervention's details will be clear (and that is fine). For example, the funding document, a recent monitoring report, and several interviews may all suggest that the intervention aims to address troubled youth; however, the details are not consistent regarding exactly what is being done, or specifically for whom (e.g., the term *troubled youth* is not defined). Thus the nutshell description provides "enough" of an understanding about the intervention to enable the monitoring or evaluation process to move forward. A clearer, more detailed intervention description will iteratively emerge through the evaluative journey.

Identifying gaps, inconsistencies, or conflicting descriptions generates *clarifying questions* to be asked during the monitoring or evaluation process (questions the evaluator asks to clarify the intervention); the answers to these questions will then inform further interviews, document reviews, or observation. Exposing inconsistencies between stakeholders' perspectives and experiences can be particularly important to engaging with realities, as opposed to what is on paper. For instance, an evaluator could ask, "Your documents say that the intervention is aimed at troubled youth. However, my interview notes suggest that you only work with youth who have drug addictions. Can you tell me a bit more about the beneficiaries?" These kinds of details affect how and what to monitor and evaluate. For instance, both interventions may be aimed at youth who regularly skip school and have failing marks; however, one intervention would also address drug addictions, while the other intervention would not. Each intervention then suggests different kinds of results (some of which are likely to be overlapping), which will affect, for example, decisions about what to monitor and evaluate.

Depending on the timing of the evaluation process, the nutshell description may describe what the intervention is designed to achieve, what is happening, what did happen, or a mix of all three (i.e., some parts have been implemented while others have not). What follows is an example of a useful nutshell description for the girls' education intervention. This example is followed throughout the next several chapters.

Girls' Education Intervention: Nutshell Description

The intervention started 3 years ago and will be funded for 2 more years. It focuses on helping school-age girls access education that is freely given in their community. The intervention mostly focuses on convincing fathers who currently prevent their daughters from attending to support their daughters to get an education. Local knowledge suggests that fathers or father figures often prevent girls from attending school. The intervention also addresses some contextual issues, such as the safety of the girls as they go to and from school. The intervention's theory is based on literature drawn from education, development, and gender studies, as well as local cultural knowledge.

The purpose of the nutshell description is simply to provide a place to *start* a more in-depth conversation about the intervention. With the nutshell description in hand, we now seek to understand *why* the intervention exists. The answer or answers to that question almost always offer a gold mine of information to be used during the entire evaluative process.

Introducing the Problem Statement

Now that we have a very broad understanding of what the intervention is, we (the evaluators) need to understand the *reason* for the intervention. An activity, policy, program, or intervention is put in place for at least one reason. It exists to fix, improve, or otherwise change something, for someone. This "something" is labeled the **problem statement.** There may be one problem or multiple problems. It may be a well-documented problem or a hidden one. Furthermore, different people may describe the problem differently, based on their knowledge, experience, beliefs, or position in society. And some people may not describe it as a problem at all.

Uses for a Problem Statement

Although an evaluation journey can start in multiple places, addressing the problem statement always occurs near the beginning, as it provides a critical foundation for the evaluative process (and it may be the beginning of an evaluator's engagement if he has been asked to help inform the design of an intervention). There are four key uses for a well-researched and clearly written problem statement:

- *It informs how to design an intervention.* Since the intervention aims to fix the problem, whoever designs the intervention needs to know what to fix (improve, enhance, influence, or change).
- *It identifies how the intervention is relevant.* If someone asks a question about the relevance of an intervention, one answer lies in the problem statement: The intervention aims to produce results that fix an identified problem. This is what makes it relevant, and it is also how potential results are identified.
- *It identifies who aims to benefit.* You are probably thinking, "Relevant to whom?" The problem statement helps to clarify who, or what, is intended to benefit.
- *It clarifies what to assess.* When an evaluator is determining what to assess in terms of progress toward results or achievement of them, information is found in the problem statement because the intervention has aimed to fix something. If the problem is fixed or is moving toward being fixed, then the intervention is likely to have been effective (here, *effective* means that the intervention did what it aimed to do).

While these four areas (how to design an intervention, how the intervention is relevant, who aims to benefit, and what to assess) are further informed through other parts of the evaluative process, a good problem statement offers a useful place to start to find some answers (see Figure 3.1).

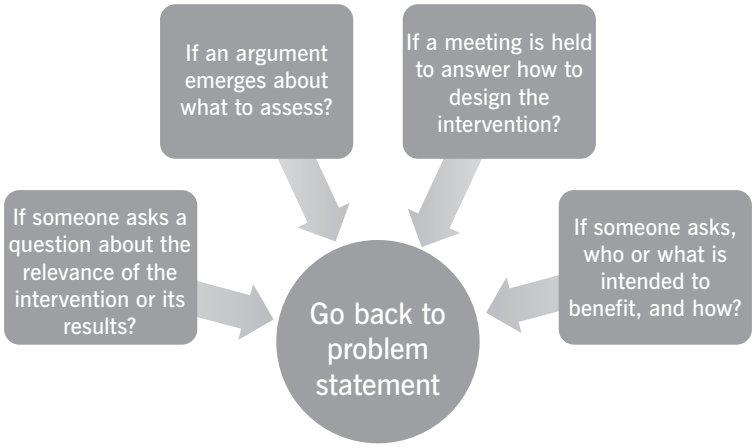


FIGURE 3.1. Problem statement: The questions it can answer.

When Labeling a Problem Statement Can Be a Problem

Sometimes a challenge to working with a problem statement is its label. Sometimes the word *problem* can be found to be abrasive or too negative. When people find the word offensive, it can become a problem in itself—mostly because when people are offended or uncomfortable, they tend to focus on the word they do not like, instead of answering the gist of a question. Thus the question does not get answered. Also, offending a person through word choice does not encourage a person to engage with you. In fact, having a person shut down in the first part of an evaluative process is likely to have a ripple effect throughout, particularly if that individual is in any way influential in the process. (To read more on language, see Chapter 13.) If the word *problem* is not appropriate, do not use it; simply substitute another word, such as *challenge*. For an evaluation to be useful, reflect constantly on how to choose words and frame questions to make the best fit with your clients, stakeholders, beneficiaries, or other actors (for more discussion of these words, jump ahead to pages 60–61). What is defined in this chapter as a *problem statement* can be relabeled to be more culturally or contextually appropriate; it is the information under the label that matters. If the label offends, replace it. Words matter.

Words Matter

How we communicate matters, what we say, and how we say it. Something we say in one situation may be appropriate, and in another it may be offensive or just confusing. Changing a word or a sentence, or an accent or emphasis on a word in a sentence, may elicit a completely different response. A colleague sent me a link to a YouTube video that provides a striking illustration of how words matter (*www.youtube.com/watch?v=Wgi0t2ap-us&feature=youtu.be*).

Identifying the Problem Statement: The Questions to Ask

Now that we know what makes the problem statement useful, and now that we realize we can call it something else if that is helpful, let's talk about how to identify one. An evaluator can ask one or more of the following questions, which are all variations of the same question, to initiate the problem statement conversation:

- What is the problem that the intervention aims to address?
- What does the intervention intend to fix, change, or improve?
- Tell me about the reasons the intervention was developed.
- What gap (or gaps) was this intervention intended to fill?

BOX 3.2. Answering Questions or Asking Them?

Answering evaluation questions is often what evaluators are asked to do—yet knowing how, what, when, where, why, and with whom to ask good questions is an essential part of being an evaluator. Throughout the book, I provide examples of questions that can be used to elicit information. However, these questions will need to be modified, adapted, or otherwise cultivated to support each unique evaluation journey. For example, you may need to use words that may be more specific or more appropriate to the context, or to change the wording of a question from present to past (or past to present) tense. Or you may need to change the common noun (e.g., *intervention*) to the most appropriate one (e.g., *program, organization, policy, activity*).

These questions can be explored by using data-gathering techniques similar to the ones used to generate the nutshell description: document reviews, interviews, observation, or all three. If the questions are used to guide an interview, consider carefully how each question is worded. These questions are written as open-ended questions, meaning that they are intended to encourage wide-ranging responses. Unless you have narrowed down the answers needed, or you only require brief answers, avoid asking closed-ended questions. At this stage in the process, you will want to encourage an in-depth discussion about the problem statement because there is so much to learn about one, and from one.

Sometimes the four questions listed on the previous page may not elicit useful responses (i.e., data). If this happens, consider one (if not more) of the following options, which take different approaches to identifying the same information.

- Your organization/intervention exists because . . . ? (Rather than have someone answer a question, have the person complete a sentence.)
- Can you tell me a bit more about the background of this intervention? (Asking for a general history often encourages a nice, easy dialogue.)
- In the next 5 years, what do you expect to see change or improve, or just be different, because of the intervention? (A time frame sometimes helps people focus.)
- Tell me a bit more about what makes the intervention critical or necessary. Who thinks this? What makes them think that? (A more informal and general conversational approach may be helpful.)
- What do those of you involved in the intervention want to accomplish? (The person states what they want to achieve. You then take the answer, and restate the flip side or opposite of that, and it becomes the problem. I demonstrate this "reverse" tactic a bit later in this chapter.)

These questions can be mixed with the first set of questions; each set is not exclusive. At this point in the process, a problem of any size can be identified; this is the starting point of the conversation. These questions can be supplemented with probes as needed, such as "What makes you think that?" or "Can you tell me a bit more?"

Different-Sized Problems: Breaking Down the Problem Statement

The problem statement can (and should) be separated into two categories: the **grand problem(s)** and **pocket problems.** The words are not intended as labels per se. Rather, they are organizing concepts to guide a discussion that will allow an evaluator to explore how an intervention addresses *what* problem, which then leads to *what* specific, intended result. Eventually, the process will link the problem statement to the intervention to the results, all at the same levels (i.e., a grand problem is linked to a grand result; a pocket problem is linked to a pocket result). I describe the process in an iterative manner in Chapters 4–10. Here we are starting to lay the groundwork.

The Grand Problem

The grand problem is the larger problem, such as gender violence, pollution, or obesity. Sometimes it can be described at a lower level or with more specificity, such as "Children have poor nutrition in Town G." In other words, the grand problem is the largest problem the intervention is aiming to address, put a dent in, or use as the "guiding light" (the light that all can see, yet realize that it may not be reached in the foreseeable future).

The Pocket Problems

All the pocket problems, when combined, form the grand problem; the reverse is also true (i.e., the grand problem can be broken down into the pocket problems). Figure 3.2 shows an example. The pocket problems can be broken down further into mini-sized ones if we continue to ask the never-ending "why" questions (e.g., in the Figure 3.2 example, "Why do few healthy food choices exist in the elementary school?"). Essentially, the problem statement needs to be broken down as far as needed to inform these three kinds of understanding:

- *Before implementation.* To understand how to design the intervention so that it addresses specific problems, and in doing so to identify where expected results will happen, and for whom.

- *After implementation.* To understand why the intervention did what it did, where, and for whom.
- *After implementation.* To understand why the intervention had (or did not have) the results it did, where, and for whom.

Additional information gathered throughout the evaluative process will also inform these understandings. In addition, how an evaluator engages with and uses the problem statement will vary, depending on when she enters the process: either forward-looking (planning an evaluable intervention), in the moment (as when supporting innovation), or backward-looking (evaluation).

BOX 3.3. Using Pocket and Grand Problems

When people respond to initial questioning about the problem statement with a grand problem, ask questions to break the grand problem down into pocket problems. If they respond with a pocket problem, build the pocket problems up to a grand problem. This is an iterative process, and the discussion is likely to go back and forth between the two. In the process, gaps may be identified, such as when the pocket problems do not build up to the grand problem, or there is no awareness of what pocket problems form the grand one. These are all part of the evaluative findings.

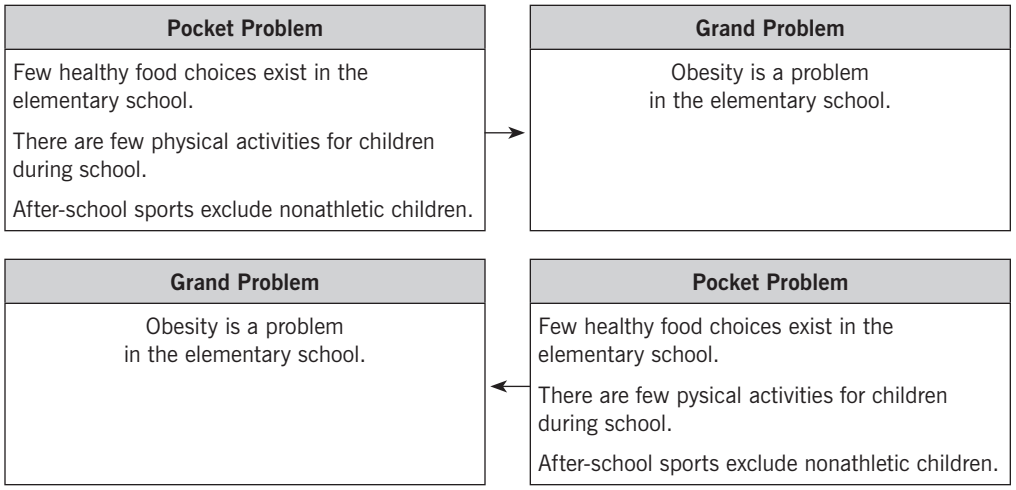
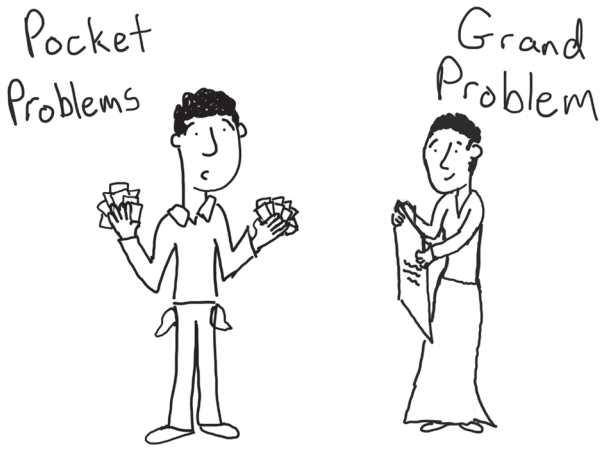


FIGURE 3.2. Pocket and grand problems.

There are no concrete definitions of grand and pocket problems, or even specific ways to write them; the terms are not intended to constitute a formula or to be used to squish problems into a predetermined check/tick box. Rather, what is offered is a loosely structured way of thinking. The thinking allows a person to conceptualize, explore, and facilitate a detailed discussion that organizes thoughts on how, and which, problems are addressed by what parts of an intervention, and how all this relates to intended (and/or actual) results. If this approach seems a bit messy, or if it is not clear how it is used in practice, skip ahead to Chapter 6 to see the concepts in action—then come right back.



A warning is in order at this point: Do not rush to finalize the problem statement. The fluid evaluation process systematically, iteratively, and repeatedly interrogates it, and in doing so, identifies the pocket and grand problems, which creates a strong foundation for any type of planning, monitoring, or evaluation. To refer to Aesop's fable of the Tortoise and the Hare (see "A Conversation with My Readers" at the start of the book), the evaluator's aim is to be the Tortoise, not the Hare—that is, to win the race with slow, steady effort. With that in mind, let's explore the problem statement from a slightly different perspective (which may overlap with the one we have just discussed, or may bring new insights): separating what causes the problem from what is a symptom of that cause. Knowing the difference, and then clearly identifying which (if not both) the intervention aims to address, will influence expected results and help determine what to assess. Curious? Read on.

BOX 3.4. A Problem Statement, Revisited

An intervention addresses something that someone considers a challenge, a problem, or something that needs improvement, fixing, or change. A statement summarizing that "something" is called a *problem statement,* which is made up of *grand* and *pocket problems*. We evaluate interventions to see whether, or to what extent, how, and for whom the

intervention has addressed (or is currently addressing) the entire problem statement, or just elements of it.

Clarifying Different Kinds of Problems, and Why It Matters

We have just noted that the problem statement has two parts: grand and pocket problems. Once those are established (or sometimes as they are being established), another perspective provides useful information: identifying whether the problem is a root cause or a symptom of that cause.

Distinguishing the Root Cause versus a Symptom

Here is an example of a root cause versus a symptom. Sometimes I have a headache. I take some aspirin, and the headache goes away. Maybe it never comes back. Maybe it does. Maybe the headache is a sign of a lack of caffeine, which is the sign of another problem (drinking copious amounts of caffeinated coffee or black tea and caffeinated diet soda can cause a caffeine addiction). Or maybe the headache is a sign that I need stronger reading glasses. Or maybe it is a sign of something more serious.

Let's say that I have another headache. Someone gives me one bottle of aspirin. When I take my aspirin, I have no headache. If I were asked at that point in time, "Does the intervention work?", my answer would be a gleeful yes. Problem solved. Or is it? When I run out of aspirin, my headache returns. Sometimes treating the symptoms (removing the pain) will not have longer-term intended results because it has not addressed the real reason (root cause) for the pain. Furthermore, the headache may result from multiple root causes, such as a need for glasses *and* a lack of caffeine. For now, let's keep it simple and assume one root cause; then I will complicate it.

Let's say the person who gives the aspirin to me digs a little deeper about the cause of my headache before he provides an intervention. Perhaps he then identifies the root cause (caffeine addiction), and provides an intervention to cure my caffeine addiction. During the intervention, the headaches grow worse (they need to get worse before they get better). If there is an evaluation at that time, I probably will provide negative feedback. And furthermore, if the funding is stopped at that point (e.g., because the evaluation says the intervention is not working), I will still have headaches, and likely my caffeine addiction as well. However, let's now say that when the intervention is fully implemented, it solves my addiction (I now drink herbal tea) and achieves my intended longer-term results of not having headaches. Now let's add back in the more complicated situation mentioned above—that the headaches are results of multiple root causes (not only a caffeine addiction, but also a need for glasses). And this is where evaluation becomes a bit more complicated. There may

be no result from my drinking herbal tea and breaking my addiction because another root cause needs to be addressed at the same time (the root problem being reading without proper glasses). Problems can be simple to address, or they can be complicated, and that is why they need to be broken down to be understood.

Assumption Infestations

Beware: Discussions that clarify root causes and symptoms can be infested with assumptions. See Chapter 4 for an in-depth look at assumptions and facts and their importance to the entire evaluation process.

There are two lessons in the headache example.

1. When an evaluator is informing the design of an intervention (forward-looking) or conducting an evaluation (backward-looking) she needs to clarify what the intervention is aimed to change, fix, or otherwise address: root cause(s), symptom(s), or both.
2. The evaluator needs to identify at what point the evaluation happens during the intervention's implementation. This second lesson is only foreshadowed here; it is discussed more fully in Chapter 6.

Why is knowing that the root cause has not been identified, or maybe just not addressed, or that only symptoms have been addressed, important for evaluators?

Identifying the root cause and symptoms can be useful to an evaluation process for two reasons:

- *The discussion of root cause and symptoms clarifies the intended results.* Knowing which problem the intervention is aimed to address clarifies the expected results. As such, it provides insight (not often *the* answer, but insight) into what to monitor and evaluate in order to understand if results happened, or why intended results did not happen.
- *It provides insight into questions on sustainability. Sustainability* is a "suitcase word" of note (i.e., it needs to be unpacked). For more on suitcase words, their dangers, and how to mitigate them, skip ahead to pages 138–140, and we dig into the concept of sustainability in Chapter 7. The key point here is that the discussion about root causes and symptoms generates a wealth of information that may support at least part of an explanation as to why something is sustained (it addresses a root cause) or is not sustained (it addresses a symptom).

Looking for the root causes and symptoms of a problem, and which of these the intervention aims (or aimed) to address, should be on your radar screen.

Avoiding the "Why" Question by Asking a "What" Question

Starting a question with "Why" can be difficult to answer, as it is essentially asking a theoretical question. Switching the "Why" to a "What" allows for a more concrete descriptive answer. Paula Bilinsky, an evaluator who introduced me to evaluation, once told me, "It is often better to handle a 'why' question with a 'what' question." For example, if you want to know why a person is attending a program, it can be challenging for many to answer the question "Why are you here today?" Asking the person, "What made you choose to come to this program today?" is likely to result in more concrete answers, as it is a more direct question.

Why—I mean, what are some examples of what happens when an evaluator identifies that the intervention is not addressing, or has not addressed, the root cause?

Consider these three evaluation scenarios in which the evaluator identifies that the intervention has not addressed the root cause, and notice how that influences evaluation use.

1. *Evaluation finding is used.* The implementer is aware that the initial research used to design the intervention was poorly done. Although the symptom was addressed, and results have been achieved, the implementer is concerned about the potential of the result lasting (the sustainability of the effect). The implementer welcomes the evaluator and his finding that the intervention has not addressed the root cause, and uses that information to make substantive changes to the intervention. Thus the evaluation is used to influence the intervention.
2. *Evaluation finding is rejected.* The program implementer has effectively addressed a symptom that is of critical importance to the donor, and has achieved the intended results. The evaluation, however, shows that the root cause was not addressed, and therefore there is a high probability that the results will not last. Both the implementer and the donor reject that evaluation finding and claim success. No changes are made to the intervention.
3. *Evaluation finding is quietly ignored.* The implementer has been under tremendous political pressure to "get going and get results." The evaluation finding that the symptoms are being addressed but not the root cause is of little interest to the implementer, who is dealing with continuing pressure to keep moving ahead. The report is shelved.

Even when an evaluator understands the importance of addressing the root cause, or at the very least of knowing that the root cause is not being addressed, evaluation clients may not bring the same appreciation. It happens. At the same time, the evaluation may be very much appreciated, and provide useful information that is used to inform the program.

Our intervention has been implemented extremely well, but the results are dismal. Could something be wrong with the problem statement?

Absolutely. There could be many reasons why the results are dismal, which should be explored (see Chapters 4–8 and 12–13 for additional potential reasons). However, for now, let's look at some common ways a problem statement may contribute to poor or unexpected results.

- *The problem statement only addresses one key root cause, when several exist.* See the headache example starting on page 51. If only one key root cause is addressed, but not the others, there may be no, or few, intended results identified.
- *One or more pocket problems did not exist when the intervention was designed.* Another reason is that at least one new pocket problem has arisen since the intervention was designed. This is often quite likely, considering that communities, people, politics, policies, or other **variables** (i.e., things that can change) may be constantly changing. For example, while childhood obesity may have been a problem 5 years ago and remains a problem today, the pocket problems, or the factors causing obesity in children, may now be different. In other words, what was *not* a problem 5 years ago may be a problem today, which then influences the intervention's ability to achieve the intended results (i.e., it is likely the wrong intervention).

It seems like asking questions about the intervention and developing the problem statement are good places to start an evaluative journey. However, these are not the questions I am usually asked at the start. In fact, no one has ever asked me to sort out the problem statement. My clients usually ask me to "just" assess results and the effectiveness of their intervention. Can you explain a bit more how to convince my clients that I need to clarify the problem statement?

Even when your clients, boss, or colleagues ask you to start in a different place, such as assessing results or identifying indicators (indicators are explained in Chapter 9), an evaluator first needs to have a basic understanding of the intervention and of what problem it aims to address. Rushing forward and trying to assess results, and linking those results to the intervention without first understanding how the intervention is linked to a problem, are likely to be challenging if not impossible. After acknowledging your clients' needs, explain

the following logic (you may want to jump to Chapter 4 to read about facts and assumptions before using the following example):

A problem statement describes what needs to be fixed, changed, or improved (it is a fact). An intervention is implemented to fix it (it is an assumption, sometimes with a few facts mixed in). If the intervention works, it will have intended results (assumptions) and perhaps unintended ones. The results should erase the problem statement, or at least put a dent in it. Thus, to understand if the intervention is effective, we first need to understand what the problem is, how the intervention aimed to influence that, and what results were expected. Understanding each of these three items informs assessment decisions—and, indeed, a monitoring framework and evaluation design.

A follow-up to that suggestion is to select one or several of the questions listed earlier in this chapter (see page 47) and engage in a discussion. For example, "Tell me a bit more about what makes the intervention critical or necessary. Who thinks this? What makes them think that?" The responses to those questions may start the necessary conversation that leads to exploring the need for a clear problem statement.

I would be remiss not to mention other real-world possibilities for engaging in problem statement discussions. Challenges may arise when an intervention has been aimed to fulfill a political or personal agenda, "fit" within what the donor wants to do, or spend down a budget, and has never been intended to address a fact-based social problem. (A problem statement needs to be fact-based, not an assumption of a fact. As noted above, to read about the need for fact-based problem statements, go to Chapter 4.)

Wait. What if I find out that the reason for the intervention is to fulfill a political or personal agenda, fit what a donor wants to do, or spend down a budget, and I need to conduct an evaluation? What happens then?

This is a tricky situation, that's for sure. Let me provide some guidance. First, tread carefully. Second, just being aware is beneficial, as it helps you to navigate the situation. Third, consider how the evaluation *process* may be useful to someone, at some point, in some way. This is called process use. To understand more about and engage with evaluation process use, see page 229. Fourth, realize that even if the intervention was to fulfill a political agenda, for example, it may still be an intervention that aimed to do some good for something or someone, and would benefit from evaluation; or maybe not.

Admittedly, the proverbial monkey wrench is thrown into the process when the real reason for the intervention is based purely on a political or personal agenda. Even so, engaging in reality, and knowing that other reasons can exist, can bring useful insight to the evaluative process. Here is an example.

A small town (let's call it Town A) has well-maintained primary schools that have consistently good academic results. Town A is affluent. The adjacent town, Town B,

lies within the same political zone (i.e., the same elected officials represent both areas). However, Town B schools have consistently poor primary school results, specifically in math and science. Most families in Town B live at or just below the poverty line. The government provides a free math and science tutoring program in Town B. Residents of Town A hear about the program and demand the same intervention for their town, where most children score average or above average on the same exams. The officials, who are concerned with re-election, decide to provide Town A with the same academic program. An evaluator is asked to evaluate the intervention (with no knowledge of these circumstances). While he finds no significant academic improvement in Town A, in Town B he identifies substantial increases in math and science scores. He may conclude that the intervention has worked in Town B but not in Town A, or that it is not necessary in Town A, and recommend that the intervention only receive further funding for Town B. Uh-oh.

Imagine, however, that the evaluator stops and asks this question: What problem did the intervention intend to address? He finds out that in Town A, the intervention has been provided to solve a political problem, and in Town B, it has been aimed to address an educational one. What *needs* to be evaluated is if Town A residents have been pleased to receive the intervention, and if so, whether they have voted for their local representative to serve another term. After all, in Town A that has been the intervention's actual goal. What is assessed in Town B should remain the same: measuring the math and science achievement of targeted children. In the real world, it is highly unlikely that an evaluation will assess how people from Town A vote, and link that to the intervention; however, knowing that alternative reasons exist for an intervention to exist (as opposed to an actual social problem) will keep an evaluator sane when he exhausts all other logical reasons, or his recommendations are rejected. (Cut funding from Town A? Um. No.) All I can say is that it is the real world out there, people, and not sweeping reality under the rug is one of the many tough aspects of being a good evaluator.

Who does the problem identification, and who does the rest of the evaluative process—the researcher, evaluator, program manager, or someone else?

Here is a very nice example of how a researcher, evaluator, and program manager can work together in harmony, which elaborates on our discussion in Chapters 1 and 2. The researcher (or the evaluator, using her research skills) identifies the problem that needs to be addressed through using a research design that provides facts; the program manager reviews these facts and designs the intervention to address one or all of the problems, noting specific results for certain people, animals, or the environment. The evaluator works with the program manager to ensure that the intervention is evaluable, and identifies the criteria that will be used to determine the intervention's value, merit, and worth. In rare cases, the three roles may be performed by one person. In more common scenarios, each role is carried out by multiple

people. Note, however, that the evaluator often arrives after the problem is identified, and often after the initial intervention is designed (though not always).

I thought that the place to start engaging in an evaluation process was to focus on the intervention's theory. Can I start there instead?

Sometimes people consider the starting place of an evaluative journey to be discussing the program's theory. I find that a more difficult and less tangible place to start. The main problem is that it *is* theory—which often does not provide a concrete enough starting place for the evaluator or for those with whom she needs to engage. Rather than discussing program theory first, I focus on untangling the logic (which is what we are doing here in this chapter, starting with the problem statement), which then provides a scaffolding for a move to a theory discussion (the process is still iterative, though; discussion on theory is just brought in later). To learn more about (and be able to facilitate a concise discussion of) theory and logic, or if you would like to start an evaluation journey by discussing intervention or program theory, please jump ahead to Chapter 8. For now, let's look at the evaluator's role in facilitating and describing the problem statement.

The Evaluator's Role

The evaluator can fill many different roles with regard to the problem statement. If she is there when the intervention is conceptualized, she can play an active role in identifying and confirming that there is indeed a problem. If she arrives after the intervention is designed or implemented, she can then facilitate a process to identify the grand problem and then unpack it into pocket problems, or the reverse. Sometimes the evaluator must sort out what the intervention is aiming to address on her own, or with her team, often through a document review. Regardless of the process, remember that an iterative interplay exists among the discussions of the problem statement, intervention, and results, and that each discussion helps to clarify and refine the others. While facilitation skills prove useful in this process, logical thinking and patience are necessities.

BOX 3.5. Facilitation

If you are using this book to learn about working in the field of evaluation, or to guide self-reflection, then facilitation skills are not yet needed. If you are using this book to engage others in an evaluative process or teach others about being an evaluator, then facilitation becomes a critical skill. Facilitation skills are developed over time, and are not something that one can learn effectively in an afternoon, in one course, or by reading a book. Strong facilitation skills demand a combination of reading, practice in a variety of settings, and self-reflection.

Learning to Be a Facilitator

If you know someone who is a good facilitator, ask to observe him in action. If you know more than one facilitator, this is even better, as different people have different facilitation styles, processes, and games. Observe, practice, reflect, observe, and practice again. If you are not fortunate enough to know at least one good facilitator, think back to a well-facilitated class, workshop, or seminar: What made the facilitation strong? What did not go well? How would you do it differently? Then practice, reflect, and practice some more.

Facilitation

- ***Have a minute?*** Take a look at the website of the International Association of Facilitators (*www.iaf-world.org/site/index.php*), which provides various types of resources.
- ***Have several hours?*** If your interest is deeper, consider reading *Facilitating Evaluation: Principles in Practice* by Michael Quinn Patton (2018a).
- ***Want to have some fun?*** Check out Chris Lysy's blog (*https://freshspectrum.com/blog*), where he provides cartoons depicting everyday evaluation conversations that bring touches of levity to facilitating evaluation processes.

The world is messy. Where I work, defining what my intervention is attempting to change is complicated. I see problems (and other influences) all over the place that likely affect what my intervention is trying to address and achieve. What about those other problems and influences? How do I sort this out, in terms of monitoring and evaluation?

The world indeed seems *simple* at times, *complicated* at other times, and *complex* at still others. These terms are associated with the Cynefin framework, which includes two more categories, *chaos* and *disorder* (Kurz & Snowden, 2003). Using these words brings in language that specifically acknowledges systems thinking in evaluation, and thus addresses the issue raised in your first statement: The world is a messy place. The evaluative process described in this book provides one basic starting place for understanding an intervention, what it aims to achieve, and what can influence it. The process slowly builds, and eventually different elements that complicate the world around us are added and bring the needed level of complexity to the discussion. We will get there, all in good time. When faced with a complex or complicated problem, I often feel overwhelmed (which throws my brain into chaos and disorder), and thus I break down the evaluative process so that it is graspable (before I have a breakdown myself). I begin just by trying to simplify, clarify, and focus. And then, and only then, do I start to bring in the complexities that I need to engage with, in order to understand what else may be influencing the program and its results.

I have just mentioned the usefulness of the Cynefin framework. I recommend an article by Snowden and Boone (2007) on applying the Cynefin framework, where they provide

real-life examples of its usefulness. Two other great thinkers in the systems field are Donella Meadows (*http://donellameadows.org*), whose approach to systems is user-friendly and easily accessible, and Bob Williams (*www.bobwilliams.co.nz*).

The perception of a problem can be just as much of a problem as the problem itself, right?

The perception of a problem is indeed a problem to be discussed. However, exploring people's perceptions is different from engaging with what they perceive to be a problem. Clear as mud? If there is a fact-based problem that girls do not go to school because they do not have access to school uniforms, then one intervention would be to provide the uniforms. However, if that is only a perception (it is not actually true, people just think that), then providing the uniforms will not solve the problem (because the intervention is addressing a perception of a problem). In this case, an evaluator needs to find out why that perception exists and address that perception. An evaluator needs to know how, or to what extent, that perception has the potential to negatively affect the intervention or its results.

Is a situational analysis the same the thing as identifying a problem statement, or a root cause and symptoms?

A situational analysis is often broader than simply identifying a problem statement. A *situational analysis* is a process that looks at a situation to understand what is happening, why, and what needs to be addressed. The analysis helps to ensure an intervention's relevance. There are several elements to a situational analysis:

- Defining the extent of the problem in that context.
- Identifying the perceptions and experiences of key stakeholders in relation to the problem.
- Identifying what already exists to address the problem, and the gaps in those already existing strategies and/or interventions.
- Identifying possible partners.

Ask if a situational analysis was done at any point in time, as it has the potential to provide a wealth of data for any evaluative process.

Wait, what about a needs assessment? What is that?

A needs assessment identifies what is needed by a specific group of people, is used in many professions, and can be conducted via many different methods. For example, a needs assessment can be used to identify what specific knowledge and skills exist, and what skills and knowledge are needed, thus identifying the gaps where training needs to take place. Needs

assessments are also conducted in a broader sense, for determining what the needs are, for example, in a community. See below for further reading.

BOX 3.6. The Reasons for Developing a Clear Problem Statement

Without an exact understanding of what an intervention is trying to change, address, fix, or make better, it is nearly impossible to be clear about anything else in the evaluative process. A clear problem statement, and the facts that support the statement, are vital for designing and focusing a relevant intervention, for assessing its progress toward or achievement of results (or lack thereof), and for understanding who is intended to benefit.

We now have a *general* idea of what the intervention is (nutshell description) and a general understanding of why the intervention exists (problem statement). We now need to understand who (person) or what (place or thing) is intended to benefit or in some way receive services from the intervention.

Useful Tools for Identifying Root Causes

- ***Have a few minutes?*** Check out this summary of the Five Whys tool (*www.betterevaluation.org/en/evaluation-options/five_whys*).
- ***Have a few hours?*** Check out what is called the *fishbone* or *herringbone* analysis, among other names (*www.mindtools.com/pages/article/newTMC_03.htm*).
- ***Have a day?*** Use one of these two websites to work through an example; each one provides guidance in a step-by-step process (*http://web.mit.edu/urbanupgrading/upgrading/issues-tools/tools/problem-tree.html* or *www.sswm.info/content/problem-tree-analysis*).
- ***Have a few hours a night?*** If you are interested in learning more about needs assessments, check out this online publication: *A Guide to Assessing Needs: Essential Tools for Collecting Information, Making Decisions, and Achieving Development Results* by Ryan Watkins, Maurya West Meiers, and Yusra Laila Visser (2012; go to *www.openknowledge.worldbank.org/handle/10986/2231*).

Introducing Beneficiaries and Six Related Terms

Beneficiaries is a common term for the persons or groups an intervention is intended to benefit. **Stakeholders** are not necessarily beneficiaries, but can be. Stakeholders are literally those who have a "stake" in the intervention and its results; in other words, the intervention and its results matter to them, in some way, for some reason. Stakeholders may or may not benefit, directly or indirectly. *A beneficiary is always a*

stakeholder; however, a stakeholder is not always a beneficiary. Stakeholders can be powerful allies or detractors in an evaluative process, or even nonentities (although this is unlikely). A **gatekeeper** is someone who has the power to block or influence access to information, people, or places; a gatekeeper can be either a stakeholder or a person who is not obviously or directly related to the intervention or evaluation. Another related term you may hear is **audience.** An audience is anyone likely to use the evaluation. The reality on the ground is that most often beneficiaries are the least likely group to read an evaluation report and use it; often the donor, manager, or implementer will use the evaluation findings to inform the intervention, which then affects the beneficiaries, though there are definitely exceptions. Knowing the audience informs how to communicate the evaluation findings. The audience is sometimes a **client,** and the term *client* is most often referring to those who commissioned the evaluation. The person who commissioned the evaluation can be called the **commissioner.** The term **actor** can also be used, particularly in evaluations that draw on a systems approach. *Actor* is an encompassing term that includes those who are interested in or likely to influence the intervention and its findings, and it can refer to more than people, such as the economy or a policy. An evaluator needs to know whom the intervention is intended to benefit; who has a stake in, or is interested in, the evaluation process and/or its findings; and who needs what information communicated to them in what way, for what reason, when, and how. For now, let's get back to discussing beneficiaries.

FOCUSING ON THE INTENDED BENEFICIARIES

The entire reason why an intervention exists is to make something better for someone or something. Thus "Who or what is going to benefit from the intervention, and in what way?" is a pretty darn important question.

When you are reading through documents or conducting interviews, ask three questions (see Figure 3.3):

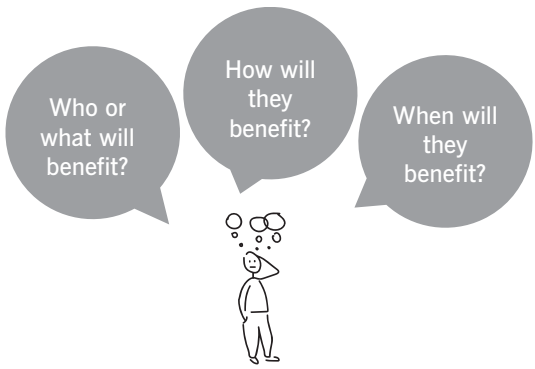


FIGURE 3.3. Considering who or what benefits.

- Who or what will benefit from the intervention?
- How will they benefit?
- When will they benefit?

The answers to these questions will provide insight into the relevance of each part of the intervention. More specifically, the answers provide information that further elaborates on the problem statement, the intervention, and the results.

- *The problem statement.* The answers to these questions identify who thinks the problem is a problem for whom. (This provides insight into who values what.)
- *The intervention.* The answers illuminate who is intended to benefit from what parts of the intervention, and how. (This conveys insight into how to design the intervention.)
- *The results.* The answers sort out who (e.g., a person or group of people) or what (e.g., a forest or an animal shelter) is intended to benefit from the intervention in what way (how does who or what benefit) and when that is intended to take place. This further clarifies the intended results (see Chapter 7 for a specific discussion of results).

Defining the Intended Beneficiaries

When evaluators ask who (or what) are the intended beneficiaries, people often tend at first to give very broad answers, listing groups or thematic areas like "children," "the environment," or "animals." It is important to probe the answer and gather as much descriptive, specific detail as possible. When "something" (or anything) is being evaluated, the more specific the "something" can be made, the clearer it becomes *what* to assess and *where* to gather information, and the lower the chances for misunderstandings become. For example, if the thematic answer is that children will benefit, probe for a bit more detail: Children who live where? What age groups? Children with siblings? Children with single parents? *These kinds of differences are important—to the beneficiaries, the design and implementation of an intervention, its intended results, how they are valued, and what to evaluate.* For this reason, the conversation should end with a rather precise (e.g., "children under the age of 12 who have unemployed parents and live under the poverty line"), and less thematic (e.g., "vulnerable children"), understanding of the beneficiary group(s). Information on the beneficiaries critically informs every part of the evaluative process.

When who or what is intended to benefit is made clear (there may be several groups of beneficiaries), there are three important follow-up questions to ask:

1. Who decided, and what process was used to decide, who are the intended beneficiaries?
2. How, if at all, were the beneficiaries involved in this decision?

3. What were the challenges or disagreements, if any, with regard to who would benefit and who would not?

These questions bring to the surface information about values, politics, and power—information that will guide what and how to monitor, how to design the evaluation, and how to interpret and value the findings. Here is an example of how engaging with potential beneficiaries influences an understanding of relevance and valuing in an evaluative process. Group R and Group G live in neighboring communities, have similar demographics (the word *demographics* means the characteristics of a group, such as economic status or age), and have received the same intervention. Group R finds the results of the intervention (how they benefitted) to be useful and important (relevant). Group G does not consider the benefits relevant; instead, it views them as nice to have, but not necessary. What can account for these two very different experiences of the intervention? It emerges that before the intervention started, several potential beneficiaries from Group R were asked what they needed or wanted, and a close version of that was provided. Potential beneficiaries from Group G were never consulted. Not being involved in any part of the decision to have an intervention, the shape of it, or its intended results may have resulted in Group G's feeling left out, alienated, or perhaps miffed; hence its members' response. Or maybe a more obvious reason is that because Group G was not consulted, perhaps the intervention did not meet Group G's needs.

Direct and Indirect Beneficiaries: The Barking Dog Example

In the evaluation process, there are typically two categories of beneficiaries: **direct** and **indirect,** sometimes called **primary** and **secondary.** An intervention should have clearly stated direct beneficiaries, and an idea of indirect or secondary ones, which may include persons, animals, places, or things (e.g., children, horses, a forest). Simply put, when a dog is barking nonstop because he is bored (the problem being that the dog is bored), and someone gives him a ball to play with, the dog is the direct beneficiary. The neighbor who has been annoyed by this barking is the secondary beneficiary—meaning that the neighbor does not benefit directly from the ball (the dog does), but they benefit in that the neighborhood is now quiet. Who is the direct beneficiary gets a bit tricky with a slight change to the orientation of the problem. Let's say that the problem is not about the dog's being bored; the problem is that two neighbors have a poor relationship. The dog's barking is just one smaller, pocket problem that has contributed to the poor relationship; the barking dog is not the *only* problem. So when the dog gets the ball and stops barking, the canine-free neighbor is now less annoyed with the dog-owning neighbor, which contributes to a better relationship. The happy, quiet dog is still a beneficiary (he receives the ball), but the neighbors are the primary beneficiaries of the quiet dog. The neighborhood is the indirect beneficiary, as it benefits when neighbors have strong relationships. Problem statements matter. Beneficiaries matter. Perspectives matter.

Once I worked with an organization that did not like the word ***beneficiary.*** *What then?*

The term *beneficiary* may be a loaded term and should not be used in situations where it brings a negative or uncomfortable meaning. I once had a similar experience where I worked with an organization in which the term *beneficiaries* suggested that a person was passive in receiving something, not an active participant in the process or service. Select a term that is culturally appropriate and acceptable to the organization with which, and the location or context where, you are working.

Eschewing Homogeneity and Examining Heterogeneity

Men are not all the same. Some are married and some are not; some are poor and some are not; and some are well educated and old, and some are young and disabled. Not all young, disabled men are the same; some are sight-impaired, while others have other disabilities. Not even all sight-impaired young men are the same: Some were born blind, some became blind through a degenerative disorder, and some became blind after an accident. A key lesson to remember is that *every difference creates a difference,* and those differences should influence the intervention's design, its implementation, its results, and the ways the results are assessed and valued. In scholastic terms, examining differences is referred to as looking at **heterogeneity** (*hetero-* means "different"); in other words, it is exploring diversity.

The next time you are in a yoga class, a supermarket, a café, or a doctor's office, look around and identify someone who initially seems like you in some basic ways, such as gender, age, height, weight, and race. Then stop and think: In how many ways is it possible that the person is different from you—ways that would make a difference to how they benefitted and valued the same service? What might they value more? Less?

While every difference makes a difference—a way of thinking attributed to Gregory Bateson (1972)—it does not mean that an intervention must necessarily be individualized for each person's circumstances (though it can be). For the evaluator of an intervention, understanding how differences influence people who were assumed to be homogeneous but were not (e.g., school-age girls, forest rangers, athletes, homeless youth) can provide explanations for why an intervention was more effective for certain people, or valued more, or valued less, by different people all assumed to be part of the "same" group. In evaluation, something is valued, and there needs to be an understanding of the source from which these values stem. While Chapter 12 covers values, and Chapters 14 and 15 talk about how different evaluation approaches involve different values, here we focus on how beneficiaries' values and perceptions of the intervention provide one of the more critical places to look for *how* to understand and value a result.

Let's go back to our girls' education intervention. The direct beneficiaries are the school-age girls who currently do not attend school. In how many ways might those girls be different, and how many of these ways would potentially influence the effectiveness of the intervention and how those girls perceive it? Some examples with the potential to make a

difference may include girls who have supportive mothers or other supportive males in the family (but not their fathers), girls who want to go to school, or girls who have brothers who go to school.

Who or What Can Be Damaged or Hurt by the Intervention?

A critical, yet often forgotten group consists of those who can be hurt, damaged, or negatively influenced by an intervention, while others (potentially or actually) benefit. To explore who or what (e.g., the rhinos, the environment) might be in this group, consider asking one, if not all three, of the following questions:

- Who or what can be harmed, hurt, upset, or damaged by the intervention?
- Who or what may have diminished power, benefits, or access to resources as a result of the intervention?
- Who thinks the intervention is damaging in some way, regardless of whether or not it is a reality (perceptions)?

Oddly, evaluation theories, models, and frameworks do not provide a specific term to describe persons, places, or things that can be potentially hurt, damaged or intentionally forgotten by the intervention. Not providing a term ignores the importance of understanding one of the more critical aspects in evaluation (and in society). As critical as it is to know who benefits, it is as critical to know who does not, and even more critical to know who is ignored, damaged, or hurt by an intervention. Perhaps we can call them the *un-beneficiaries*. The knowledge of who or what can be harmed, upset, or damaged by an intervention can broaden the insights of a program implementer or designer, which can lead to changing the intervention or mitigating potentially unpleasant situations. For the evaluator, the information regarding who or what may be un-beneficiaries can be obtained from data that explains implementation challenges, unexpected results, or negative findings, such as data on groups or persons who may have either overtly blocked the intervention or subtly sabotaged it. Engaging with the un-beneficiaries also provides a place to explore values that can be used to identify criteria for how to value findings. Perhaps the un-beneficiaries were the group that needed most to benefit and did not, or perhaps the damage done to them by achieving the results outweighs (from their perspective) the intervention's achievements. Consider, for example, the advantage of having a road link between communities. This link may provide access to better medical care, fresher produce, and jobs. At the same time, the road may damage the environment, including local plants or animals and the water supply.

Data provided by the un-beneficiaries can also explain why certain groups or individuals provide critical feedback on the intervention, even when the intervention has achieved all its intended results. Furthermore, asking the question "Who are the un-beneficiaries?" can provide necessary data to explore and understand larger questions of social justice.

Socially Just for Whom?: Some Food for Thought

Thinking about beneficiaries and un-beneficiaries often poses an unspoken challenge for an evaluator who wants to encourage a socially just society through being an evaluator or conducting an evaluation. The evaluator who concerns herself with social justice needs to grapple and come to terms with this question: In any evaluative process, what is considered socially just for whom, and who thinks that? If there are multiple perspectives on what is socially just, how does the evaluator choose to negotiate and engage with those perspectives, and how does that affect the evaluation, its findings, and its recommendations? At the same time, she may need to grapple with the challenge of implementing an evaluation that is *not* focused on social justice—for example, when evaluation users (or clients) are not particularly interested in that issue.

Beneficiaries. Indirect beneficiaries. Un-beneficiaries. Stakeholders. Clients. All these groups are likely to provide useful and varying perspectives (often useful just because they are varying). Yet an evaluator's role goes beyond asking questions and gathering data from multiple perspectives and viewpoints; she also needs to make sense of it all. At times, at least for me, the sense-making role can be a bit (OK, a lot) overwhelming. Yet making sense of data from multiple perspectives is a core reason why evaluators play a critical role in the evaluative process, and in society.

Including Beneficiaries in the Process: Placebo or Panacea?

In a discussion of beneficiaries, the word *participatory,* and the whole notion of *participation,* are often brought into the conversation. Specific evaluation approaches have different meanings and make use of different processes for participation. Some approaches that emphasize participation, yet in different ways, include democratic evaluation, empowerment evaluation, feminist evaluation, developmental evaluation, and, (no surprise) participatory evaluation. See Chapters 14 and 15 if you want to jump ahead to read about evaluation approaches. For now, a very practical discussion wraps up the chapter and takes a down-to-earth perspective on the whole concept of participation.

Irene Guijt (2014) describes participatory approaches as those that in some way involve an intervention's stakeholders (of which beneficiaries are one group, as noted on pages 60–61) at any stage of the evaluation. She notes that the reasons for choosing a participatory approach may be either pragmatic or ethical, or a combination of the two. They may be pragmatic because better evaluations are achieved (e.g., better data, better understandings of the data, more appropriate recommendations, better uptake of findings); they may be ethical because taking such an approach is viewed as the right thing to do. Given these potential benefits, participation in evaluation is often presented as a win–win situation. Nonetheless, such benefits are not a sure thing.

I want to share a stream-of-consciousness reflection on the words *participation* and *beneficiaries*. When someone says, "Oh, I am doing a participatory process," it always makes me

wonder: Who is participating? Who chooses whom to participate? What if a group or person does not want to participate (i.e., forced participation or refusal)? How does a person participate, at what level, and at what point in the process? How does it benefit the evaluation? How does it benefit the beneficiary?

When I am asked to implement a participatory process in a community, particularly in an impoverished one, I feel awkward about the fact that I am getting paid and the people participating are probably not. They might get a cup of tea, or have their transportation costs reimbursed, or receive a small honorarium. People often attend meetings, organized for an evaluation process, at a personal or economic cost to themselves; is that OK? It might be; it depends on the context. Using the word *participatory* does not automatically mean that an evaluation is an inherently fair or just process. A participatory process can be quite exclusionary, for example, allowing (or encouraging) some groups to participate and not others. Or the word can be misleading when groups are invited to participate at times when they cannot; when there are other barriers to participation; or when, even if people do participate, their voices in no way influence the evaluation or its related decisions. Then there is the realization that sometimes beneficiaries do not want to participate; should we (the evaluators) badger them until they do, so that we can say the process was participatory?

Sometimes researchers or evaluators think that gathering data from a person through an interview constitutes a participatory process; that makes me shudder. If that were so, then every evaluation that gathers primary data would be participatory. When we label a process participatory, or when we are being asked to conduct a participatory process, we need to be very clear about how the approach will be used in practice, how it is beneficial to the evaluation, and how it is intended to be useful to the people involved.

Consider asking some, if not all, of the following questions:

- How will the process be participatory, and for whom?
- How will participation be compensated, if at all?
- Whose participation is deemed useful and beneficial to the evaluation?
- Who will benefit through participating, and in what ways?
- Who can get hurt if they are, or are not, involved?
- What part of the process is participatory? Is it participatory for different people at different points in the process?
- How does a participatory approach lead to a more credible evaluation? (For a discussion of evaluation credibility, please see Chapter 5.)
- How does participation support a fair and just process?
- And lastly, who is making all of the decisions in regard to these questions, and how are they made?

Phew. It feels good to share.

Let me end this section by stating that participatory approaches, when done well, can be informative, useful, and beneficial to the participants and the evaluation. Here are some examples of how participatory processes are useful: building the capacity of those who participate (when mentoring and training are part of the process); improving the relevance and accuracy of an evaluation report; bringing the local or internal knowledge necessary to interpret findings; increasing process and evaluation use; and contributing to the face validity of an evaluation. *Face validity* means that an ordinary person can understand the evaluation questions, methods, and analysis without special methodological training, and the evaluation makes sense in a straightforward, logical way.

Another concept that is also intertwined in the conversation about beneficiaries and participation is *buy-in*. Obtaining buy-in is a very popular concept when an evaluation brings together multiple stakeholder groups. In these types of evaluative processes, an often-heard expression is "We need to get buy-in . . . through a participatory process." Be careful with the concept and what it implies. If evaluators aim to create buy-in to an assumption or a value, for example, this also likely means that someone or some group is not only dominating, but winning. The aim of an evaluation that has multiple stakeholders is not for one stakeholder to win; the focus is on gathering and understanding those multiple perspectives, and making meaning out of everything that has been heard and gathered. All stakeholders should recognize their voices (e.g., perspectives, needs, insights) in the report, and therefore find the evaluative process, its findings, and the evaluator to be credible and trustworthy.

Engaging Beneficiaries

- ***Have a few minutes?*** Take a brief look at how Derek Sivers explains that recognizing difference is useful (*www.ted.com/talks/derek_sivers_weird_or_just_different#t-142837*).
- ***Have a few hours?*** Read Irene Guijt's (2014) report, *Participatory Approaches* (Methodological Briefs—Impact Evaluation 5, published by the UNICEF Office of Research in Florence, Italy). While it focuses on impact evaluation, the document provides a wealth of information on participatory approaches, a useful glossary, and excellent resources in the bibliography.
- ***Have a few hours a night?*** Consider selecting a few chapters in *Collaborative, Participatory, and Empowerment Evaluation: Stakeholder Involvement Approaches* (Fetterman, Rodriguez-Campos, Zukoski, & Contributors, 2018).
- ***Have a few more hours a night?*** Sandra Mathison's (2005) *Encyclopedia of Evaluation* provides nice summaries of the approaches mentioned in this chapter, plus a plethora of others. Another place to read about the approaches mentioned is the Better Evaluation website (*www.betterevaluation.org*). Look under the Themes and Approaches tabs.

WRAPPING UP

Writing a nutshell description of an intervention is a useful way to start exploring what to evaluate. This description initiates the discussion of three core concepts: intervention, problem statement, and beneficiary, which are critical to any evaluative process. Identifying the interplay among them demonstrates the necessary interconnected thinking and iterative nature of the evaluation journey. The problem statement, when distinguishing between grand problems and pocket ones, and root causes and symptoms, informs the reason for an intervention, provides guidance in how to design one, and anticipates its results. When a problem statement is well thought out, it provides insight into questions often asked throughout the evaluative process, such as the relevance of the intervention and its results, who aims to benefit, and where results will take place.

Recognizing the role of different stakeholders, one group of whom are the beneficiaries, can generate an awareness of the multiple perspectives that exist; specifically, it can bring to the surface a multitude of values and perceptions that can be used to reflect on the intervention, inform an evaluation design, and influence the criteria for valuing an intervention's results. The discussions in this chapter inform the initial, iterative steps of any evaluative process, and when addressed well, provide a delightful gift to the evaluator. The next chapter delves a bit deeper into the evaluative process and asks you to consider how you know what you know, challenges you to consider how others may think very differently, and explores how to engage with thinking that perhaps does not resonate with your own.

Our Conversation: Between You and Me

The conversation with me is intended to coach you through similar situations you will encounter in the real world.

In the chapter, I shared a stream of consciousness about the concept of participation. I mentioned that sometimes people equate participation with a fair and just process. What was your reaction to that? Pondering about what participation *actually* means helps me to think through what will be my response when a client asks for a participatory process, before they ask me. I use the thought process and questions highlighted in the chapter to do my pondering. What will you do the next time someone says, "I want a participatory process?" How will you respond? What questions will you raise, if any?