3 EDITION

Michael Bamberger
Linda Mabry

RealWorld Evaluation

Working Under Budget, Time, Data, and Political Constraints

Praise for the Third Edition of *RealWorld Evaluation*

This is an invaluable resource for both novice and experienced evaluators. It contains a variety of tools and recommendations to successfully design and implement effective evaluations for any size and type of program.

—Sebastian Galindo, University of Florida

This book moves the study of evaluation from the theoretical to the practical, so that evaluators can improve their work. It deals with most of the real issues that evaluators face, particularly at the international level.

—John Mathiason, Cornell Institute for Public Affairs

RealWorld Evaluation moves forward from where other evaluation textbooks stop. *RWE* challenges the evaluator to ask the difficult questions that can impact the design, implementation, and utilization of the evaluation. *RWE* then leads the reader through how to find efficient solutions to minimize these constraints.

—Karen McDonnell, Milken Institute School of Public Health

RWE is a must-read for students of program evaluation—the framework and emphasis on practical constraints makes it an invaluable tool for learning the art and science of public policy.

—Amanda Olejarski, West Chester University

This is one of the most practical textbooks in the field of evaluation that I have encountered. Its recognition of the limitations that affect program evaluation provides students with a realistic understanding of the difficulties in conducting evaluations and how to overcome these difficulties.

—David C. Powell, California State University, Long Beach

Any research class focusing on real-world evaluation should start with this text; it is comprehensive, well-organized, well written, and thoroughly practical.

—Jeffrey S. Savage, Cornerstone University

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 600 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

RealWorld Evaluation

Third Edition

In appreciation of the support of many colleagues and workshop participants around the world who contributed to the development and testing of many of the ideas and methods included in this book . . . and to Elizabeth for her patience and encouragement throughout its writing.

—Michael

To evaluators everywhere working to improve programs and society through practice that is not only competent but also marked by sensitivity and integrity.

—Linda

RealWorld Evaluation

Working Under Budget, Time, Data, and Political Constraints

Third Edition

Michael Bamberger

Independent Consultant, Former World Bank Senior Sociologist

Linda Mabry

Washington State University

Foreword by Jim Rugh

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Copyright © 2020 by SAGE Publications, Inc.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.

Printed in the United States of America

ISBN: 9781544318783

Acquisitions Editor: Helen Salmon
Content Development Editor: Chelsea Neve
Editorial Assistant: Megan O'Heffernan
Production Editor: Jane Martinez
Copy Editor: Terri Lee Paulsen
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Scott Oney
Indexer: Joan Shapiro
Cover Designer: Candice Harman
Marketing Manager: Shari Countryman

This book is printed on acid-free paper.

19 20 21 22 23 10 9 8 7 6 5 4 3 2 1

BRIEF CONTENTS

Appendices available online at

and also in the various eBook versions of the text.

DETAILED CONTENTS

Appendices available online at

and also in the various eBook versions of the text.

LIST OF BOXES, FIGURES, AND TABLES

BOXES

FIGURES

TABLES

LIST OF APPENDICES

Appendices available online at

and also in the various eBook versions of the text.

FOREWORD

by Jim Rugh

I coauthored the 2006 and 2012 editions of the *RealWorld Evaluation* book with Michael Bamberger and Linda Mabry. However, when Sage invited us to produce a third edition, I opted out, since I am taking retirement seriously! (I'd rather hike than write!) Nevertheless, when Michael asked me to write a foreword, I agreed, since I'm still impressed by what an impact this book, and the multiple trainings on this subject we've led in more than 20 countries, have had on the global evaluation profession.

Permit me to begin with a little history. Michael and I led a workshop at the 2000 American Evaluation Association (AEA) conference in Honolulu, Hawaii. We called it "Shoestring Evaluation." Participants appreciated it for being so realistic and practical. When I shared the idea with my colleagues of the Atlanta-Area Evaluation Association (AaEA), they wisely suggested we change the name to "RealWorld Evaluation," because "Shoestring Evaluation" sounds as though the focus is just on doing evaluation as cheaply as possible. We agreed, and the RealWorld Evaluation name has become quite popular.

As mentioned above, we've led workshops on this topic in over 20 countries. Some of the various versions and translations of the PowerPoint presentation used in such workshops are accessible at https://realworldevaluation.org/.

Turning now to the book itself: We are flattered by the many professors of evaluation courses who have told us that the *RealWorld Evaluation* book is one of the basic texts they use in their courses. As they have noted, the first section of the book is a practical guide for how to conduct evaluations as rigorously as feasible, given the many constraints typically faced, and then the second part of the book provides more thorough introductions to many aspects of evaluation.

The first section can serve as a reality check for those who practice evaluation—whether evaluators themselves or clients responsible for developing the terms of reference and supervising evaluations. The second section serves as a slightly more advanced textbook for those wanting to learn more about various approaches to and dimensions of evaluation.

I am pleased that my colleagues have taken the time and effort to update and significantly add to the previous editions of the *RealWorld Evaluation* book. The field of evaluation, in the international development field as well as evaluation in many other contexts, continues to evolve. Thus, readers and those using this book as a training manual will find it a useful resource as they prepare to organize and conduct evaluations in, well, the real world.

In the Preface the authors describe what has been added in this third edition of the book, but let me make just a few comments here.

There are often discussions about the relationship between research and evaluation. Certainly, evaluators use many of the techniques of research methodologies used in the various sectors.

However, many of those research designs are, quite frankly, rather simplistic. For example, randomized control trials (RCTs) seem to focus on the "impact" of one intervention, on the assumption that all else is the same between the experimental sample and the so-called control group. In the real world, things are rarely that simple. In most contexts there are multiple influences and factors that can contribute in various ways to any discernible "impact"—which calls for evaluative perspectives that can cope with complexity. I am glad to see increased emphasis on such perspectives in this updated edition of the *RealWorld Evaluation* book.

They mention the acceptance of the Sustainable Development Goals (SDGs) by the United Nations in 2015, and how they present an additional challenge to governments and to evaluators to measure progress toward their achievement (see https://en.wikipedia.org/wiki/Sustainable_Development_Goals). Though each of the 17 SDGs has a particular focus, there are surely many factors that can contribute to (or detract from) the achievement of those goals, some within the control of governments and other development agencies. How can one take those various influencing factors into account when designing and conducting evaluations of the SDGs in any country or subregion of a country? Complexity, indeed!

Speaking of the SDGs and global trends, 2015 was the International Year of Evaluation. During that landmark year, EvalPartners launched a number of networks. (See the Networks/Task Forces tab at https://evalpartners.org/.)

I mention these networks to contend that this updated version of the *RealWorld Evaluation* book makes valuable contributions to evaluators and others who are working on any of those cutting-edge dimensions of evaluation. One of the networks, EVALSDGs, focuses on issues related to the challenges of evaluating the SDGs themselves. EvalGender addresses issues of gender equality and women's empowerment, which is discussed in the new Chapter 17 of this book. EvalYouth is very proactively encouraging young persons to become involved in the evaluation profession. It also includes "emerging evaluators"—specialists in various sectors who are becoming engaged in evaluation. Surely studying this *RealWorld Evaluation* book would be among the useful resources as they prepare for this profession.

In summary, let me suggest that whether you are a young or emerging evaluator, or, indeed, an experienced evaluator who wants to expand your understanding of a greater variety of dimensions and approaches to evaluation, this book can serve as a valuable resource for you. Let me add that persons with responsibilities, on behalf of the agencies they work for, for commissioning evaluations, would do well to read this book (at least the first section) in order to have a better understanding of the options and dynamics of what they really are looking for and can expect from an evaluation of their programs.

OK, as one of the coauthors of the first two editions of this book, I obviously am biased! But truly, in my experience during my long career as an evaluator, and one who has led countless workshops on evaluation, I have found the *RealWorld Evaluation* book to be one of the most practical yet comprehensive resource materials available. And I've heard other colleagues express similar sentiments.[1]

I'm pleased that Michael and Linda have dedicated their time, talents, and creative energies to update and significantly add to the previous versions, by producing this third edition of *RealWorld Evaluation*. I highly commend it to your reading, study, and application in the real world of evaluation!

[1]See testimonials about RWE at https://realworldevaluation.org/rwe-resource-materials.

PREFACE

There are many books and guidance notes available on how to design a methodologically rigorous evaluation for a variety of sectors and initiatives. However, there are very few, if any, written on how to realistically apply adaptations of these methodologies to real-world contexts. Very seldom does an evaluator have enough time, funds, political space, supportive organizational structures, or the capacity to implement well-designed methodologies, and consequently many end up making pragmatic compromises in order to have any evaluation at all.

This book is to help the evaluation implementors and managers make the best choices when faced with real-world constraints. Our experience from having organized workshops in more than 20 countries (see Acknowledgments) is that there is widespread interest in learning how to conduct adequately rigorous evaluations when it is not possible to follow standard research/evaluation methodologies. These obviously are questions that most evaluation practitioners and evaluation users frequently face. Although many evaluators are familiar with the *fast-and-dirty* evaluation studies that largely ignore concerns about rigor and validity of conclusions, very few have systematically addressed the challenge of how to ensure maximum methodological rigor under a particular set of real-world constraints. To the best of our knowledge, no previous evaluation textbook systematically addresses all of these questions.

This book addresses the challenges of conducting program evaluations in real-world contexts where evaluators and the agencies commissioning evaluations must adapt the evaluation to a range of constraints, pressures, and influences. These include budget and time constraints; critical data are not available or are of poor quality; the interventions are poorly designed and with essential elements unarticulated; the programs work within organizational systems that are influenced by multiple agencies and actors; and they face administrative procedures not well suited to the conduct of a rigorous evaluation. Evaluators are also subject to pressures from many sides concerning the evaluation methods to use, what should be studied (or left out), who should be involved, and how and to whom findings should be disseminated. While trying to juggle and reconcile all these factors, it is also essential for the evaluator to follow standards of professional practice and to adhere to evaluation codes of conduct.

We were inspired to write a book about RealWorld Evaluation (RWE) because it became clear from our evaluation teaching and consultancies that these topics were not systematically addressed in most evaluation textbooks or guidelines. So while it was very easy to find help on how to conduct a rigorous impact evaluation with no budget and time constraints, and no critical data missing, it was very difficult to find practical guidance on how to conduct evaluations of an acceptable level of rigor and validity while working under real-world budget, time, and data constraints—all the while reconciling different political perspectives within organizational structures prioritizing different agendas, often at odds with a rigorous evaluation.

AUTHORSHIP OF THE CHAPTERS OF THE THIRD EDITION

The first two editions of the book were coauthored by Michael Bamberger, Jim Rugh, and Linda Mabry, with the exception of Chapters 6 ("Political Constraints"), Chapter 9 ("Standards and Ethics"), and Chapter 13 ("Qualitative Evaluation Methods"), which were written by Linda Mabry for the first edition and updated for the second and third editions, and Chapter 16 ("Evaluating Complex Projects, Programs, and Policies"), which was written by Michael Bamberger for the second edition and updated for the third edition. Two new chapters are included in the third edition: Chapter 17, "Gender Evaluation," which was written by Michael Bamberger with important contributions from Maliha Khan, and Chapter 18, "Evaluation in the Age of Big Data," which was written by Michael Bamberger.

Although Jim Rugh decided not to continue as coauthor for the third edition (see Jim's foreword to this edition), he had actively contributed to most of the chapters in the first and second editions and must continue to receive credit for this.

WHAT IS NEW IN THE THIRD EDITION?

Since the publication of the second edition in 2012, the field of evaluation has continued to evolve. In addition to new ways of understanding how evaluations are affected by the contexts within which they are implemented, there are also some very significant changes in the contexts themselves. The most significant of these since the first edition in 2006 is the tremendous advancement and spread in digital technology and the concomitant advances in the ability to collect, transmit, store, access, analyze, and disseminate data and findings. Digital technology has also made a tremendous difference to ordinary people's ability to access information and services, which is a significant change to the context and has implications for the impact of initiatives that needs to be factored in. There is also a potential negative impact of the digital revolution on issues of data privacy, protection, and responsibility, which is something that the development field is slowly realizing.

While it is not possible to cover all of the new developments, we tried to incorporate some of the most relevant ones, either in new chapters or as themes that occur throughout the book:

The Sustainable Development Goals (SDGs). One of the most significant changes within the development field was the introduction in 2015 of the Sustainable Development Goals (SDGs) as the successor of the Millennium Development Goals (MDGs). The SDGs have become a focus of many of the debates on evaluation methodology and an impulse for work on new approaches such as complexity-responsive evaluation. The insistence on equity, universal application in scope, multisectoral focus, and multiplicity of actors means that the SDGs present many new challenges and opportunities for evaluation.

Gender equality and women's empowerment. Gender equality and women's empowerment (GEWE) have become more mainstream and are now recognized as key development goals by most development agencies, even some of the most conservative. However, the lag between understanding issues conceptually, creating effective implementation strategies, and finally having easily accessible tools and the capacity to evaluate them is still there, and it is widely acknowledged that many evaluations do not adequately address these goals. The new Chapter 17 focuses on how to address issues of gender equality and women's

empowerment in and through evaluation. It gives an overview of a range of approaches for assessing program impact on gender equality and women's empowerment. In the chapter, we suggest there is currently a continuum of approaches—ranging from simple gender analysis that disaggregates a set of standard evaluation indicators by sex, through more comprehensive gender analysis frameworks, to a range of feminist evaluation approaches. It is stressed that the choice of the GEWE evaluation approach involves value orientations as well as methodological considerations.

Digital technology, big data, and data science. The massive advances in digital technology and implications for the collection, storage, transmission, analysis, and dissemination of data in the field of evaluation is very significant. The biggest changes are the speed and accuracy, and consequently cost, with which primary data can now be collected, tabulated, transmitted, and analyzed. The ability to crowdsource data and to use big data and data science are coming to play an increasingly important role in program selection, design, and management. There is a new Chapter 18 devoted to this topic. Data analytics is opening up a whole new approach for the collection and analysis of data. Machine learning, predictive analytics, and the creation of integrated data platforms are just a few of the new approaches being introduced. However, at this point the evaluation community has been slower to adopt big data than have other parts of the development community (for reasons that we will discuss); but there is tremendous potential for the application of these tools in development evaluation.

With every opportunity there is also challenge. The same features of digital technology, big data, and data science that create the opportunities for evaluation also pose potential threats. With increasing amounts of data being collected, transmitted, and stored online, the threat posed by potential data breaches or misuse of sensitive and personal data is tremendous. The increasing use of automated decision making, where complex and frequently nontransparent algorithms make important operational and policy decisions with little human oversight, is beginning to raise concerns. We explore the opportunities, challenges, and ways to build bridges between data scientists and development evaluators and the need for better ethical and responsible data practices.

Complexity. With the increasing recognition that many development interventions are complex comes the use of complexity-responsive evaluation designs to capture the effects of these interventions at the policy, program, and project levels. However, at the time of writing, while there is an extensive theoretical literature on complexity, there are not many complexity-responsive evaluation methodologies accessible to evaluators who do not have a background in systems analysis and complexity science. The updated Chapter 16 presents a framework for the analysis of the four main dimensions of complexity and a checklist for rating the level of complexity of a project or program on each dimension. It also includes a five-step complexity-responsive evaluation strategy that combines unpacking a complex program into a set of components, each of which can be evaluated separately using conventional evaluation tools, with the application of complexity science tools that respect the special characteristics and challenges of complex programs.

Emergence. Closely related to complexity is the recognition that programs and the environments in which they operate are constantly changing. Consequently, how a program is organized, the challenges it faces, and the results it achieves may be very different after three or five years than they were when the program was designed and its outputs, outcomes, and impacts were initially defined. Many conventional evaluation designs still continue to compare results with the original baseline and are not well suited to capturing these dynamic processes. This discussion will also incorporate ideas from the evolving field of adaptive management.

Focus on agency. Realist evaluation examines how programs are modified in response to interactions with the target populations and with all stakeholder groups. It focuses on the agency of affected groups rather than viewing them as passive recipients of the services offered by development agencies. It also recognizes that programs have different effects on different groups, at different times, and in different locations. It asks questions about who benefits, when, how, and why. Realist evaluation is also an effective approach for understanding complexity and emergence.

Negotiating boundaries. All evaluations make explicit or implicit decisions on what is to be included and what is left out—in essence, what is valued and what is not. The boundary definitions affect, among other things, the definition of program goals, who are considered to be the beneficiaries and affected groups, the outcomes that are valued and assessed, and the time horizons over which effects are measured. Critical to the boundary definitions are questions of power and control: who decides how the boundaries are defined, who controls the evaluation resources, and who decides who is consulted and who receives the evaluation report. Often the decisions are made by a small group of powerful stakeholders and frequently many of the boundary decisions are implicit and never formally stated or discussed. Discussions on boundaries are an important element of complexity analysis (critical systems heuristics) as well as of evaluations focused on gender and women's empowerment, equity, and social justice (critical evaluation). Chapter 17 illustrates the importance of boundaries in gender analysis, but the approaches and issues are applicable in most kinds of evaluation.

Equity-focused evaluation. The SDG priority to ensure "no one is left behind" reflects the recognition that the benefits of most development programs are not equally distributed and that some groups are often left out or may even be worse off as a result of an intervention. The marginalized and vulnerable groups are difficult to identify, and the assessment of equity outcomes continues to be one of the weakest aspects of most evaluations. Equity issues are discussed throughout the book, particularly in the new Chapter 17.

INTENDED AUDIENCES AND READER'S GUIDE

The book is intended for four main audiences.

1. *Evaluation practitioners* who design, conduct, and advise on evaluation methodology and use.
2. The *users of evaluation*, including agencies who commission evaluations, the staff of agencies or programs being evaluated, policymakers, academics, groups affected by the programs being evaluated, advocacy groups, civil society, and the media.
3. The *managers of individual evaluations and of evaluation systems*, including evaluation offices in donor agencies and national implementing agencies, evaluation managers in consulting firms and research agencies, and national and sector agencies that plan and coordinate national or sector evaluation strategies.
4. The *university and research community* where evaluation methodology is taught and which is one of the main sources of evaluation expertise.

The following table provides a brief summary of the organization of this book. We hope this road map will help you know where you are and where you want to go as you explore this book.

A Reader's Guide to *RealWorld Evaluation* (Third Edition)		
	Preface	**Introduction to the Third Edition of This RWE Book**
Part I: The Seven Steps of the RealWorld Evaluation Approach ***Part I presents a relatively nontechnical review of each of the seven steps of the RealWorld Evaluation approach to the design, implementation, analysis, and use of evaluations. It can be used by managers and other readers who need an overview of the field, without too much technical detail, or an introduction to Part II.***		
Chapter 1	Overview: RealWorld Evaluation and the Contexts in Which It Is Used	A preview of the seven steps and the basics of how the RWE tools and techniques can help evaluators and their clients cope with a variety of constraints typically faced when conducting evaluations in the real world.
Chapter 2	First Clarify the Purpose: Scoping the Evaluation	This discusses the varying expectations of different evaluation clients and the use of program theory to help articulate expectations for the evaluation and to identify the key outputs and outcomes to be assessed. The chapter addresses the importance of a contextual analysis to understand the environment in which the program operates, and to identify potential constraints on how the evaluation can be conducted.
Chapter 3	Not Enough Money: Addressing Budget Constraints	Strategies for reducing the costs of an evaluation are given. New material is included on potential applications of big data for reducing costs and time of data collection and difficult-to-access data.
Chapter 4	Not Enough Time: Addressing Scheduling and Other Time Constraints	Strategies are presented for addressing two kinds of time constraints: resource constraints that limit the number of person weeks that can be covered within the budget and ways to reduce the elapsed time between the start of the evaluation and the delivery of the final report.
Chapter 5	Critical Information Is Missing or Difficult to Collect: Addressing Data Constraints	Strategies are presented for addressing different kinds of data constraints due to missing or difficult-to-collect data: reconstructing baseline data for retrospective evaluations and collecting information on difficult-to-reach groups or on sensitive topics.
Chapter 6	Political Constraints	Strategies are discussed for addressing power dynamics and political issues that affect the design and scope of the evaluation and the collection and use of data at different stages of the evaluation. The trade-offs are also discussed between the public's "right to know" and the client's desire to restrict access to the evaluation process and findings.
Chapter 7	Strengthening the Evaluation Design and the Validity of the Conclusions	A framework is presented for the identification and assessment of different kinds of threats to the findings and recommendations of quantitative, qualitative, and mixed-method evaluations. A set of templates are included in annexes for assessing threats to validity and for presenting the threats to clients.
Chapter 8	Making It Useful: Helping Clients and Other Stakeholders Utilize the Evaluation	The chapter defines "useful" and "influential," identifies reasons why evaluations are not used/influential, and proposes guidelines to help increase utilization.
Part II: A Review of Evaluation Methods and Approaches and Their Application in *RealWorld Evaluation* ***For those who would like to dig deeper***		
Chapter 9	Standards and Ethics	This chapter outlines the basic "good practice" principles and standards that guide the evaluation profession.

(Continued)

(Continued)

Chapter 10	Theory-Based Evaluation and Theory of Change	The chapter discusses the increasing use of theory-based evaluation, particularly theory of change, in most kinds of evaluation. We cover both the simpler kinds of models that can be used as a framework for project evaluations and the more complicated models required for multiagency, multicomponent program evaluations. The main criticisms of theory-based evaluations are also reviewed.
Chapter 11	Evaluation Designs: The RWE Strategy for Selecting the Appropriate Evaluation Design to Respond to the Purpose and Context of Each Evaluation	This chapter reviews the stages in the selection of an evaluation design and presents a wide range of experimental, quasi-experimental, and nonexperimental designs. Statistical rigor is only one of several dimensions of a methodologically sound design, and strong statistical designs are often weak on other important dimensions, while there are many situations in which nonexperimental designs are the best option. Appendix 11 presents case studies illustrating how each of the 19 designs has been used in the field.
Chapter 12	Quantitative Evaluation Methods	These chapters review and contrast the key elements of quantitative and qualitative approaches to evaluation. Chapter 14 shows how mixed-method designs can combine the strengths of both approaches. Utilizing mixed methods is an integrated strategy involving unique approaches at each stage of the evaluation.
Chapter 13	Qualitative Evaluation Methods	
Chapter 14	Mixed-Method Evaluation	
Chapter 15	Sampling Strategies for RealWorld Evaluation	Reviews approaches to sample design for quantitative and qualitative approaches as well as mixed-method evaluations. The role of statistical power and effect size in estimating sample size is explained.
Chapter 16	Evaluating Complex Projects, Programs, and Policies	Discusses the move toward more complex, multicomponent, national-level development programs and the demands that this creates for new complexity-responsive evaluation designs. A range of promising new approaches are discussed, and a five-step complexity-responsive evaluation strategy is presented. This combines the techniques of complexity science with an "unpacking" strategy that breaks complex programs into a set of components, each of which can be evaluated separately using familiar, conventional evaluation tools.
Chapter 17	Gender Evaluation: Integrating Gender Analysis Into Evaluations	In this chapter we propose gender-responsive evaluation approaches to strengthen the analysis of the differential impacts of development programs on women and men. A continuum of approaches are presented, ranging from simple sex disaggregation of a set of commonly used indictors, through more comprehensive gender analysis frameworks, to some of the most widely used feminist evaluation approaches. It is stressed that the choice of evaluation approach is based upon values and ideology as well as methodological and organizational considerations.
Chapter 18	Evaluation in the Age of Big Data	This chapter reviews the wide range of big data recollection and analysis methods that can potentially be applied in development evaluations. The chapter also discusses the reasons why big data is less integrated into development evaluation compared to other areas of development such as research and in the design and management of development programs, particularly in areas such as health, education, and emergency relief. Strategies are proposed for building bridges between data scientists and development evaluators.

Part III: Managing Evaluations		
For readers involved with the funding and management of evaluations		
Chapter 19	Managing Evaluations	This chapter defines and discusses the main stages in the design, management, dissemination, and use of an evaluation at the agency, sectoral, and national levels.
Chapter 20	The Road Ahead	Our final comments on some of the main approaches and issues discussed in the book are presented in this final chapter.
Online Appendix		
More than 250 pages of appendices, providing supporting material for the chapters, are available at **study.sagepub.com/bamberger3e** and also in the eBook versions of the text. Additionally, on the website can be found instructor and student resources for download, designed especially for this edition of the text to help with instruction and comprehension.		

ONLINE RESOURCES

For accompanying student and instructor resources designed to assist your understanding of the book, visit the NEW dedicated companion website at **study.sagepub.com/bamberger3e**. Over 250 pages of appendices are available there for download, detailing the examples covered in the text with figures, tables, and case studies.

Instructor resources include:

- Chapter-specific slide presentations highlighting key points from the text, ideal for use in lectures and review.
- Lecture notes which summarize and outline each chapter, providing structure for class lectures.
- Figures and tables from the book available for download.

Student resources include:

- SAGE Research Methods case studies, carefully selected to draw connections to real-world examples of research concepts and methods covered in the text.

ACKNOWLEDGMENTS

SAGE and the authors gratefully acknowledge the feedback from the following reviewers during the development of the third edition:

Lisa Boyce, *Utah State University*

Tamerin A. Capellino, *Brandman University*

Sebastian Galindo, *University of Florida*

Muriel Harris, *University of Louisville*

John Mathiason, *Cornell Institute for Public Affairs*

Karen McDonnell, *Milken Institute School of Public Health*

Amanda Olejarski, *West Chester University*

John E. Paul, *University of North Carolina at Chapel Hill*

David C. Powell, *California State University Long Beach*

Michael L. Rowland, *University of Louisville*

Lillian C. Wichinsky, *University of Arkansas at Little Rock*

The earlier editions of the book benefited greatly from the advice and encouragement of Lucia Fort, Raymond Gervais, Michael Hendricks, Patrick Leung, Michael Patton, Mike Ponder, Michael Scriven, and Brett S. Sharp. Special thanks to Professor Gary Miron and his graduate students at the Western Michigan Evaluation Center for their detailed critiques of the earlier edition. We also thank Karen Ginsberg, Barbara Levine, Linda Morra, and Ray Rist at the International Program for Development Evaluation Training (IPDET); Dr. Sivagnanasothy, Mallika Samaranayake, and Soma de Silva of the Sri Lankan Evaluation Association; Patrick Grasso, Keith MacKay, Markus Goldberg, Manuel Castro, and Maurya West Meiers at the World Bank; Nobuko Fujita at the Foundation for Advanced Studies in International Development in Tokyo; Mary Cole at the Development Bank for Southern Africa; Zenda Ofir of the African Evaluation Association; Alexey Kuzmin of the International Program Evaluation Network; Esteban Tapella and Pablo Rodríguez Bilella of ReLAC; Marcia Paterno Joppert of Rede Brasileira de Monitoramento e Avaliação (Brazilian M&E Network); Anzel Schonfeldt and Benita Van-Wyk of the South African Monitoring and Evaluation Association; Ram Chandra Khanal and Ramesh Tuladhar of the South Asian Community of Evaluators; Beate Bull and her colleagues at NORAD; Verena Knippel and her colleagues at Sida; Sophie Punte and her colleagues at ESCAP; Cate Rogers and her colleagues at AusAID; and the Atlanta-area Evaluation Association, among others, for offering the opportunity to organize workshops and seminars on Shoestring Evaluation and RealWorld Evaluation as the approach developed over the past 15 years. We also greatly acknowledge the constructive feedback we received from colleagues during our RWE workshops in many countries, including Bangladesh, Brazil, Canada, Costa Rica, the Czech Republic, Egypt, Greece, India, Indonesia, the Ivory Coast, Japan, Nepal, the Netherlands,

Norway, the Philippines, Portugal, the Republic of Ireland, Russia, Sierra Leone, South Africa, Sri Lanka, Sweden, Ukraine, and the United Kingdom, in addition to multiple AEA conferences in the United States.

Chapter 16 ("Evaluating Complex Projects, Programs, and Policies") benefitted greatly from lessons learned during workshops organized on complexity-responsive evaluation with Mita Marra, Frans Leeuw, Jos Vaessen and Estelle Raimondo at the International Program for Development Evaluation Training (IPDET) and the European Evaluation Society, at the African Evaluation Association with Benita Williams, and at the American Evaluation Association and the Latin American Research Network (RELAC) with Jeneen Garcia, Aaron Zazueta, and Anupam Anand (Global Environment Facility). Similarly, Chapter 18 ("Evaluation in the Age of Big Data") benefited from workshops organized at the American Evaluation Association and at MERL Tech with Pete York and Kerry Bruce, and at the Development Café in Jakarta with Rick Davies and Valentine Gandhi.

Several chapters of the third edition also benefited from valuable comments received from Martha Arranz and Simone Lombardini (OxFam UK), Shravanti Reddy (UN Women), Jane Whynot (University of Ottawa), Gustav Petersson (Swedish Research Council), and Swapnil Shekhar (Sambodhi Consultants). Special thanks to Maliha Khan, who made major contributions to the gender evaluation chapter.

ABOUT THE AUTHORS

Michael Bamberger has been involved in development evaluation for 50 years. Beginning in Latin America where he worked in urban community development and evaluation for over a decade, he became interested in the coping strategies of low-income communities, how they were affected by and how they influenced development efforts. Most evaluation research fails to capture these survival strategies, frequently underestimating the resilience of these communities—particularly women and female-headed households. During 20 years with the World Bank he worked as monitoring and evaluation adviser for the Urban Development Department, evaluation training coordinator with the Economic Development Department, and senior sociologist in the Gender and Development Department. After retiring from the Bank in 2001, he has worked as a development evaluation consultant with more than 10 UN agencies as well as development banks, bilateral development agencies, NGOs, and foundations. From 2001–2016 he was on the faculty of the International Program for Development Evaluation Training (IPDET), and from 2001–2011 he was on the Foundation for Advanced Studies in International Development (FASID) in Tokyo. Recent publications include (with Marco Segone) *How to Design and Manage Equity Focused Evaluations* (2011); (with Jim Rugh and Linda Mabry) *RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints* (2012, 2nd ed.); *Engendering Monitoring and Evaluation* (2013); (with Linda Raftree) *Emerging Opportunities: Monitoring and Evaluation in a Tech-Enabled World* (2014); (with Marco Segone and Shravanti Reddy) *How to Integrate Gender Equality and Social Equity in National Evaluation Policies and Systems* (2014); (with Jos Vaessen and Estelle Raimondo) *Dealing With Complexity in Development Evaluation* (2016); and *Integrating Big Data Into Development Evaluation* (2017).

Linda Mabry is a faculty member at Washington State University, specializing in program evaluation, student assessment, and research and evaluation methodology. She currently serves as president of the Oregon Program Evaluation Network and on the editorial board for *Studies in Educational Evaluation*. She has served in a variety of leadership positions for the American Evaluation Association, including the Board of Directors, chair of the Task Force on Educational Accountability, and chair of the Theories of Evaluation topical interest group. She has also served on the Board of Trustees for the National Center for the Improvement of Educational Assessments and on the Performance Assessment Review Board of New York. She has conducted evaluations for the U.S. Department of Education, National Science Foundation, National Endowment for the Arts, the Jacob Javits Foundation, Hewlett-Packard Corporation, Ameritech Corporation, ATT-Comcast Corporation, the New York City Fund for Public Education, the Chicago Arts Partnerships in Education, the Chicago Teachers Academy of Mathematics and Science, and a variety of university, state, and school agencies. She has published in a number of scholarly journals and written several books, including *Evaluation and the Postmodern Dilemma* (1997) and *Portfolios Plus: A Critical Guide to Performance Assessment* (1999).

PART I

THE SEVEN STEPS OF THE REALWORLD EVALUATION APPROACH

OVERVIEW

RealWorld Evaluation and the Contexts in Which It Is Used

The chapter begins with an overview of the RealWorld Evaluation (RWE) approach, the contexts in which RWEs are conducted, and the many different constraints, pressures, and influences under which evaluations are formulated, conducted, disseminated, and used. The **RWE** approach was originally developed to address four of the most common constraints evaluators face: budget, time, and data constraints, and political influences. Subsequently, participants in workshops highlighted the importance of issues concerning organizational structures, and management and administrative arrangements. Two common RWE scenarios are reviewed. The first is when the evaluator is brought in at the start of the **program**[1] but with constraints on the types of information that can be collected or the designs that can be used. The second concerns retrospective evaluations where the evaluator is not called in until the program is nearing completion. For most retrospective evaluations, no baseline data have been collected and no comparison (control) group has been identified. The widespread use of retrospective evaluations is due to the fact that this is an approach used by the interdependent evaluation offices that form part of most multilateral development banks and many UN and bilateral development agencies.

1. WELCOME TO REALWORLD EVALUATION

Most evaluators are familiar with situations in which programs are nearing completion before clients begin to think seriously about evaluating whether the programs are achieving their objectives and producing the intended **impacts**. Usually, the belated interest in evaluation is motivated by the need for solid **evidence**[2] on which to base decisions about whether the program should be continued or perhaps expanded. When the evaluations do finally get underway, many have to be conducted under budget and time constraints, often with limited access to baseline data and comparison groups. Consequently, it is difficult to apply the most rigorous evaluation designs.

[1]Unless there is a need to be more specific, we use the term "program" to refer to any kind of intervention at the project, program, sector, national, or international level.

[2]Bold technical terms are defined in the Glossary at the end of this book.

Although more resources are often allocated to evaluation in developed countries, many evaluators in the United States, Canada, Europe, Japan, and Australasia report that they operate under similar constraints to those faced by their colleagues in developing countries.[3] As if these problems were not enough, many evaluations are often conducted in *political* environments in which funding agencies, clients, and key stakeholders have strongly held views on what the "right" evaluation methods should be, what types and amounts of information should be collected, and which groups should and should not be asked to comment on (or even see) the **findings**. New evaluators soon discover that seemingly straightforward "technical" issues—such as whether to use randomized selection of program and control groups; the choice of qualitative, quantitative, or mixed-method designs; and whom to interview and what questions to ask—can provoke strong reactions from clients and stakeholders.

Despite the difficult circumstances under which many evaluations have to be conducted, there is a growing demand from funding agencies, governments, and civil society for systematic impact evaluations, including whether the program met its objectives and should be continued or expanded to other communities or locations. Consequently, there is a strong demand from many sides for evaluators to answer questions such as those proposed by Stern et al. (2012):

- To what extent can a specific net impact be attributed to the intervention?
- Did the intervention make a difference?
- How has the intervention made a difference?
- Will the intervention work elsewhere?

Many evaluators also address the more detailed questions proposed by Realist Evaluation (Pawson, 2013; Pawson & Tilley, 1997) concerning: who benefits? where? when? how? and why? There is also an increasing awareness that, in order to be considered credible, evaluation conclusions need to be supported by sound evidence and not just opinions—although there are often major disagreements as to what constitutes credible evidence.[4]

The RWE approach presented in this book was developed in response to the demand for guidance on how to conduct sound evaluations when faced by these kinds of constraints, accommodating organizational structures and administrative procedures, while at the same time ensuring maximum possible methodological rigor within the particular evaluation context.

RWE is based on the following seven-step approach, summarized more specifically in Figure 1.1 and described in detail in Chapters 2 through 8:

- Step 1: *Planning and scoping the evaluation.* Before selecting the evaluation design, it is important to fully understand the purpose of the evaluation, the information needs and expectations of the clients and stakeholders, and the constraints and pressures under which

[3]One of our colleagues who has worked with major U.S. foundations that support community-level initiatives stated that there is a huge unmet need in the United States for material on how to conduct evaluations when working with very limited financial and professional resources. He stated that his "and other foundations make lots of small grants. There is often not enough money in the grants to hire an external consultant. And the recipients of these small grants don't have the capacity to do internal evaluation. The evaluation work done by these nonprofits is usually pretty bad. I don't really know of any materials targeted to this group."

[4]For two publications on the question of credible evidence, see Donaldson, S., Christie, C., & Mark, M. (2009). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage; and Rieper, O., Leeuw, F., & Ling, T. (2010). *The evidence book: Concepts, generation and use of evidence.* New Brunswick, NJ: Transaction.

they are working. What is the client's bottom line? What do different stakeholders really want from the evaluation, and how will the results be used? Difficult choices have to be made to accommodate budget and time constraints or to recognize the limitations of the available data. This step also includes getting agreement on the articulation of the **program theory model/ logic model** (see Chapter 10), which in addition to clarifying the underlying model on which the program is based, also helps identify the critical hypothesis and linkages in the program implementation model on which the limited evaluation resources should focus. It is also essential to identify and spell out key assumptions on which the program design is based. Often program management and program designers have deeply held beliefs about how and why the program will work, but often these are not spelled out in program documents.

Evaluators have come to recognize the importance of working with stakeholders to define the *boundaries* of the program and consequently of the evaluation. Is the goal of the program to produce benefits for a clearly defined group of program beneficiaries, or is it intended to also affect a broader population (other families in the program communities, neighboring communities, or wider population groups)? Is it only intended to achieve a limited number of clearly defined outcomes, or is it hoped that the program will contribute to broader outcomes, perhaps over a longer period of time? How these boundaries are defined will have an important effect on program design and potential impacts and also on how the evaluation is designed. The clarification of boundaries will have an important effect on the evaluation. The situation will often arise where the evaluators believe that the program could have broader outcomes (both positive and possibly negative) than those defined in the program results framework. It is essential for evaluators to reach a clear understanding with program management on whether it is possible to assess outcomes that are broader than those in the results framework and the program design. Often management only want evaluators to assess the defined program objectives, so it is essential for the scope to be clarified at the start of the evaluation. As we will see in Chapter 17, the issue of boundaries is often quite sensitive in gender-focused evaluations.

The scoping phase also involves identifying and assessing evaluation design options that are feasible within the cost, time, and data constraints that a particular evaluation will face, followed by assessing the strengths and weaknesses (i.e., threats to validity and adequacy) of each option. The different design options are then discussed with clients, emphasizing the trade-offs involved in each option, and an agreement is then reached on which design would be most feasible and acceptable to the client. (We get into more detailed coverage of evaluation designs in Chapter 11.) While the debate on whether there is a "best" evaluation design continues, the steady increase in the use of mixed and multiple methods designs has resulted in a recognition of the benefits of combining quantitative and qualitative designs (discussed in Chapter 14). It is also important to recognize that there are at least six widely used evaluation designs, each of which has strengths and weaknesses for addressing different kinds of evaluation questions. We argue that there is no single "best" evaluation design, and that the choice of design will be determined by the questions being addressed, the real-world constraints, the understanding of contextual factors, and the methodological preferences of the different stakeholders.

A final issue concerns whether the program should be considered "complex" and if so, whether a complexity-responsive evaluation design is required. The evaluation of complex programs will usually require the use of more expensive and methodologically rigorous evaluation designs (see Chapter 16).

Finally, the scoping phase must also understand the broader political, economic, sociocultural, historical, and environmental context within which the program will be implemented. Strategies for contextual analysis are discussed in Chapter 16.

- Step 2: *Strategies for addressing budget and other resource constraints.* How many evaluators have been told by the client, "We really need a rigorous and professional evaluation as it is important to assess impacts, but . . . unfortunately our budget has been cut." Step 2 describes options for reducing costs. These include simplifying the evaluation design, reducing the amount of data to be collected, making greater use of secondary data, revising the sample design and sample size, and streamlining data collection and analysis. In addition to budget, it is also important to assess organizational constraints, technical and other human resource constraints, and increasingly constraints on access to new information technology capacity.

The rapid emergence of new information technology[5] over the past few years means that one option for reducing costs (and time) may be to incorporate smartphones and big data into the collection and analysis of evaluation data (see Chapter 18).

- Step 3: *Strategies for addressing time constraints.* In addition to many of the approaches used in Step 2, strategies include planning ahead to avoid delays and bottlenecks, particularly during the short periods when outside consultants are in the field; building impact-related indicators into routine program monitoring data collection; and using videoconferencing to reduce travel and to permit more frequent interactions between the evaluation team and agency staff. ICT can also reduce the amount of time required for data collection and analysis.

- Step 4: *Strategies for addressing data constraints.* These include addressing problems concerning the lack of important data or data quality when the evaluation is not commissioned until late in the program cycle. Thus, one needs to consider a number of approaches for *reconstructing* baseline data. These include using secondary data sources, recall, key informants, focus groups, construct mapping, and participatory group techniques such as PRA (participatory rural appraisal). Chapter 4 covers techniques for collecting information on sensitive topics and on difficult-to-reach groups. As these groups are often the poorest and most vulnerable, there will often be pressures to ignore sensitive questions and inaccessible groups.

BOX 1.1

The Implications of New Information Technology for RealWorld Evaluation

New information technology (NIT) comprises new technologies for the collection, analysis, and dissemination of data. It combines big data, Information and Communication Technology (ICT), and the Internet of Things (IOT). Big data sources include, but are not limited to, satellite images and remote sensors, digital financial transactions such as ATMs, telephone call records and purchases of airtime, call-in radio programs, social media such as Twitter and Facebook, administrative records, and survey databases from government agencies, donor agencies, and nongovernmental organizations (NGOs). ICT data comes from

(Continued)

[5]New information technology (NIT) covers handheld devices such as smartphones and tablets and wearable and remote sensors (all of these are usually referred to as Information and Communication Technology [ICT]) and big data, such as satellite images; data streams from Twitter, Facebook, and other social media; phone records and electronic financial transfers; and audio and video recordings. In addition to the huge volume and speed with which the data is collected, and the need for advanced computing facilities for its analysis, most big data shares the characteristic that it was generated remotely for a purpose other than the evaluation or research for which it can be used. New information technology is discussed in Chapter 18.

(Continued)

mobile phones and other handheld devices, and IOTs include devices for recording health, travel, and sensors attached to devices such as refrigerators or domestic and public electrical supply systems. However, these huge data sources are of limited practical utility without the use of smart data analytics that transform big data into user-friendly applications such as interactive maps and charts for data visualization, the identification of patterns and associations, and prediction. As we will emphasize throughout the book, NIT has major implications for development evaluation as it makes possible (i) the economical collection of much greater amounts of data, (ii) the collection of many new kinds of information that were not previously possible, (iii) the use of new forms of data analytics and predictive modeling, and (iv) the visualization of large and complex data sets in an easy to access and understand format. Furthermore, the data often can be collected, analyzed, and disseminated in almost real time.

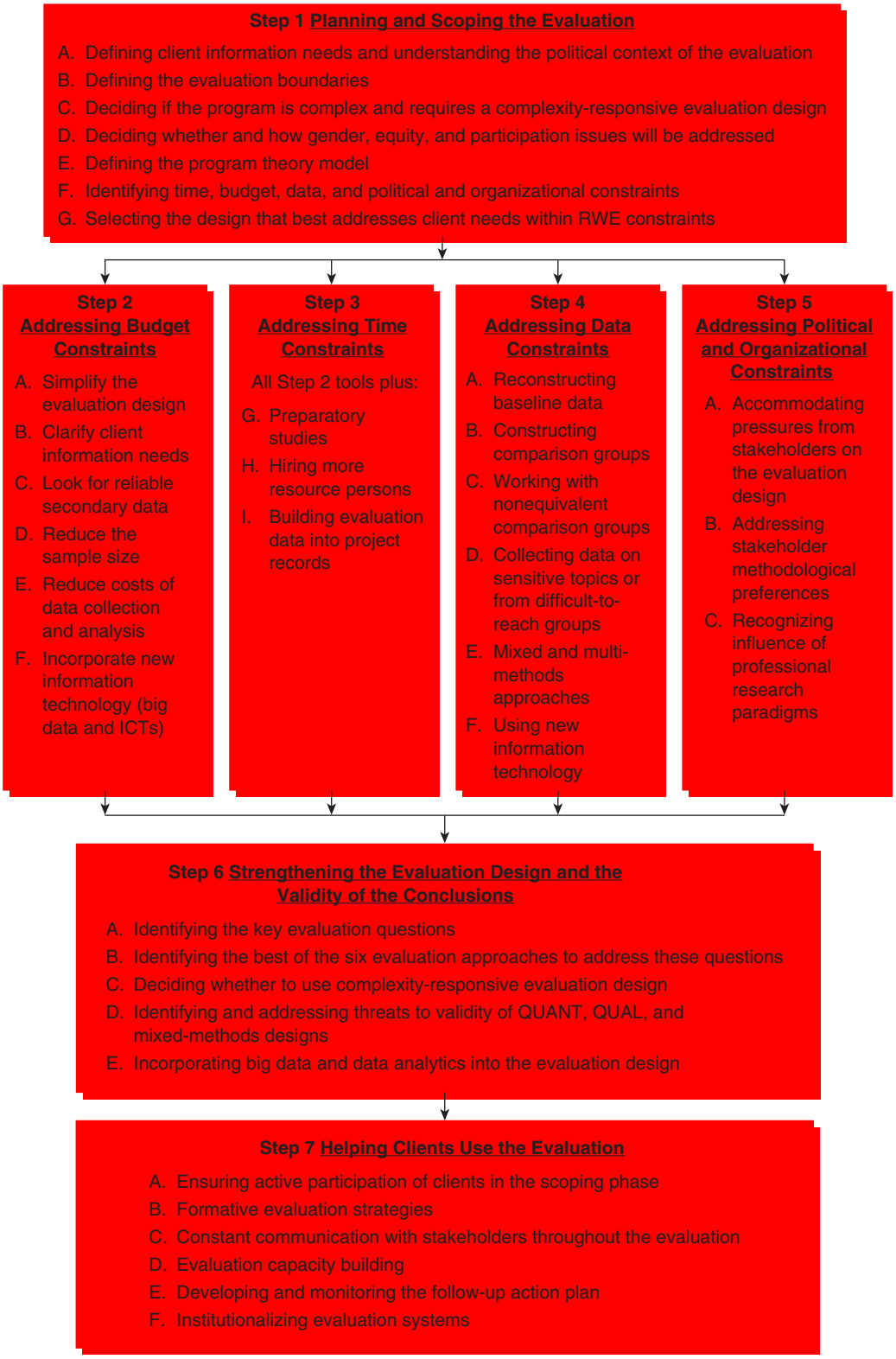
However, as we will continually emphasize, NIT also brings challenges and potential risks for both evaluators and policymakers. For example, most big data was collected for a different purpose, and it is often analyzed through proprietary algorithms; therefore, it is often difficult to surmise the quality or meaning of the data or to detect potential biases (e.g., ethical or political) in how it is used and the social consequences of these uses. There are also organizational and institutional challenges, as coordination between data analysts and evaluators is often quite weak, with the result that much of the analysis and interpretation is done by data scientists who are often not familiar with conventional evaluation approaches, and who use analytical methods with which most evaluators are still not very familiar.

Big data offers tremendous opportunities for the future directions of evaluation, but it also presents many challenges.

Since the publication of the second edition, there have been rapid advances in the fields of new information technology, making it possible to collect and analyze data much faster and more economically (see Box 1.1 and Chapter 18). There are also issues concerning whether the new methods of collecting and analyzing data are promoting more participatory and *inclusive* evaluation approaches, giving greater voice to poor and vulnerable groups, or whether they may lead to more *extractive* approaches, whereby information is collected from and about vulnerable groups, often without their knowledge, and where decisions are made about development programs and policies without any consultation with the affected groups.

- Step 5: *Understanding and coping with political and organizational factors influencing how the evaluation is designed, implemented, disseminated, or used.* It is important to identify the key actors and their political perspectives and how these affect their orientation to the evaluation. We identify political issues arising at the outset of an evaluation, during implementation, reporting, and use of the evaluation, and we propose strategies for addressing all these issues. There are also important professional and ethical issues concerning who should be given information on the evaluation and when. Often, the client would like to limit who sees and comments on the evaluation draft, whereas the evaluator may feel that the report should be given to the mass media and to the different stakeholder groups potentially affected by the program. We will return to these ethical issues throughout the book. The design, implementation, and use of evaluations are also affected by institutional and organizational constraints, often referred to as *the political economy of evaluation*. As programs become larger and more complex, these factors can play an increasingly important role in how effectively an evaluation can be conducted. Issues of coordination or competition among multiple agencies and departments can constrain the kinds of information that can be collected and how evaluation findings are disseminated and used. The ongoing discussions on how to evaluate the Sustainable Development Goals (SDGs) illustrates the challenges of conducting multicomponent, multisectoral evaluations in over 100 countries with multiple stakeholders at the international, regional, national, and local levels.

FIGURE 1.1 ● The RealWorld Evaluation Approach



- Step 6: *Strengthening the evaluation design and the validity of the conclusions*. We argue there is no single "best" evaluation design, and the choice of design should be determined by the questions of concern to stakeholders, the real-world context within which the evaluation will be implemented, and the findings used. We identify six sets of widely used evaluation designs, plus three additional designs that are applicable to complexity-responsive evaluations (see Chapter 16). Each of these designs has areas of application and its respective strengths and weaknesses. Given the widespread popularity of experimental designs (the most common of which is the randomized control trial), we discuss the great importance of the experimental approach to evaluation, as well as the many practical and methodological limitations of RCTs (randomized control trials) in real-word contexts. We also discuss the need to understand *threats to validity*, how these affect the findings and recommendations of the evaluation, and how the threats can be addressed once they have been identified. Appendices 7.1 through 7.4 present worksheets for assessing the validity of QUANT, QUAL, and mixed-method designs, and for communicating the evaluation findings and recommended follow-up to managers and policymakers.

- Step 7: *Helping clients use the evaluation.* It is important to ensure clients and other key stakeholders are actively involved from the start and that they "buy into" the evaluation; maintain contact with clients throughout the evaluation and ensure that by the time the major reports are published they do not contain any surprises for the client; and adapt the presentation of findings to the preferred communication style of different stakeholders. On a broader level, this also involves helping institutionalize evaluation systems at the sector and national level (see Chapter 19). Smartphones and data analytics now offer attractive ways to easily present findings (e.g., data visualization) and to reach wider audiences.

2. THE REALWORLD EVALUATION CONTEXT

The RWE approach was developed to assist evaluators in both developing and developed countries to conduct evaluations with budget, time, data, political, and organizational constraints. In one common scenario, the client delays contracting an evaluator until late in the program when the funding agency (government, international development agency, foundation, etc.) is about to decide whether to continue to support a program or possibly launch a larger second phase. Such tardiness occurs even when evaluation was built into the original program agreement (see Box 1.2). With the decision point approaching, the funding agency may suddenly realize that it does not have solid information on which to base a decision about future funding of the program, or the program-implementing agency may realize it does not have the evidence needed to support its claim that the program is achieving its objectives. An evaluator called in at this point may be told it is essential to conduct the evaluation by a certain date and to produce "rigorous" findings regarding program impact although, unfortunately, no comparable baseline data are available.

In other scenarios, the evaluator may be called in early to help develop the monitoring and evaluation (M&E) plan but may find that for budget, political, or methodological reasons it will not be possible to collect data on a comparison group to determine program impact by comparing participants with nonparticipants (a **counterfactual**). In some cases, it may not even be possible to collect baseline data on the program population for purposes of analyzing progress or impacts over time. Data constraints may also result from difficulties of collecting information on sensitive topics such as HIV/AIDS, domestic violence, post-conflict reconstruction, or illegal economic activities (e.g., commercial sex workers, narcotics, or political corruption).

Determining the most appropriate evaluation design under these kinds of circumstances can be a complicated juggling act involving trade-offs between available resources and acceptable standards of evaluation practice. Often, the client's concerns are more about budgets and deadlines, and basic principles of sound evaluation design may receive a lower priority. Box 1.2 illustrates this difficult trade-off between budgets and deadlines on the one hand and desired standards of methodological rigor on the other. Failure to reach satisfactory resolution of these trade-offs may also contribute to a much-lamented problem: low use of evaluation results (see Chelimsky, 1994; Operations Evaluation Department, 2004 and 2005; Patton, 1997). RWE is a response to the all-too-real difficulties in the practical world of evaluation.

BOX 1.2

A Familiar Evaluation Story

When a social development fund was launched in an African country a few years ago, it was suggested that a baseline study be conducted as the first phase of a longitudinal impact evaluation study. The program manager asked, "What is the point of spending money and time on a baseline study when we do not know if the program model will work in our country?" He also indicated that staff members were under pressure to launch the program and could not spend time on something that would not be useful until the program was completed. Three years later, when the possibility of a second program was being discussed, consultants were called in to conduct an impact evaluation study. It was agreed that it was unfortunate that no baseline data were available to permit a rigorous measurement of the changes produced by the program. The consultants had to try to reconstruct baseline data using methods described in Chapter 5.

3. THE FOUR TYPES OF CONSTRAINTS ADDRESSED BY THE REALWORLD APPROACH

Table 1.1 illustrates the different ways in which RWE constraints interact in the contexts in which evaluations are conducted. In some cases, the evaluator faces a single constraint. For example, the budget may be limited but there is plenty of time. Or the evaluation may begin at the start of the program with no time constraint, but the evaluator is told that for political or ethical reasons it will not be possible to collect data on a comparison group. However, many unlucky evaluators find themselves simultaneously contending with several—or all—of these constraints!

3.1 Budget and Other Resource Constraints

Sometimes funds for the evaluation were not included in the original program budget, and the evaluation must be conducted with a much smaller budget than would normally be allocated. As a result, it may not be possible to collect the desired data or to reconstruct baseline or comparison group data. Lack of funds may create or exacerbate time constraints because evaluators may not be able to spend as much time in the field as they consider necessary. Box 1.3 makes the point that it is important to understand whether the main constraint is budget or time (or both), because the best strategy will often be different in each case.

TABLE 1.1 ⬢ RealWorld Evaluation Scenarios: Conducting Impact Evaluations With Time, Budget, Data, and Political and Organizational Constraints

The constraints under which the evaluation must be conducted				
Time	**Budget**	**Data**	**Political and Organizational**	**Typical Evaluation Scenarios**
X				The evaluator is called in late in the program and told that the evaluation must be completed by a certain date so that it can be used in a decision-making process or contribute to a report. The budget may be adequate, but it may be difficult to collect or analyze survey data within the time frame.
	X			The evaluation is allocated only a small budget, but there is not necessarily excessive time pressure. However, it will be difficult to collect sample survey data because of the limited budget.
		X		The evaluator is not called in until the program is well advanced. Consequently, no baseline survey has been conducted either on the program population or on a comparison group. The evaluation does have an adequate scope, either to analyze existing household survey data or to collect additional data. In some cases, the intended program impacts may also concern changes in sensitive areas, such as domestic violence, community conflict, women's empowerment, community leadership styles, or corruption, on which it is difficult to collect reliable data even when time and budget are not constraints.
			X	The funding agency or a government regulatory body has requirements concerning acceptable evaluation methods. For example, in the United States, the No Child Left Behind Act of 2001 includes funding preference for certain types of research designs. In other cases, a client or funding agency may specifically request qualitative data, tests of statistical significance regarding measured program effects, or both.
			X	There is overwhelming indication that the evaluation is being commissioned for political purposes. For example, an evaluation of the effects of conservation policy might be commissioned to stall its expansion.
			X	There is reason to suspect that the evaluation will be used for political purposes other than or contrary to those articulated in preliminary discussions. For example, an evaluator might suspect that an evaluation of charter schools might be used (and even misused) by a client with known advocacy for privatization of education.
			X	Problems of coordination or even rivalries among agencies with different objectives, implementation strategies, and monitoring systems often affect the design, implementation, and utilization of the evaluation.
X	X			The evaluator has to operate under time pressure and with a limited budget. Secondary survey data may be available, but there is little time or few resources to analyze it.
X		X		The evaluator has little time and no access to baseline data or a comparison group. Funds are available to collect additional data, but the survey design is constrained by the tight deadlines.
	X	X		The evaluator is called in late and has no access to baseline data or comparison groups. The budget is limited, but time is not a constraint.
X	X	X		The evaluator is called in late, is given a limited budget, and has no access to baseline survey data; no comparison group has been identified.

Note: To simplify the table, the possible combinations of political and organizational constraints with the other three factors have not been included in the table.