

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



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

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

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

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

List of Boxes, Figures, and Tables	xxiv
List of Appendices	xxxix
Foreword by Jim Rugh	xxxiii
Preface	xxxv
Acknowledgments	xlii
About the Authors	xliv

PART I • THE SEVEN STEPS OF THE REALWORLD EVALUATION APPROACH

Chapter 1	• Overview: RealWorld Evaluation and the Contexts in Which It Is Used	2
Chapter 2	• First Clarify the Purpose: Scoping the Evaluation	19
Chapter 3	• Not Enough Money: Addressing Budget Constraints	39
Chapter 4	• Not Enough Time: Addressing Scheduling and Other Time Constraints	55
Chapter 5	• Critical Information Is Missing or Difficult to Collect: Addressing Data Constraints	72
Chapter 6	• Political Constraints	94
Chapter 7	• Strengthening the Evaluation Design and the Validity of the Conclusions	105
Chapter 8	• Making It Useful: Helping Clients and Other Stakeholders Utilize the Evaluation	127

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	144
Chapter 10	• Theory-Based Evaluation and Theory of Change	153
Chapter 11	• Evaluation Designs: The RWE Strategy for Selecting the Appropriate Evaluation Design to Respond to the Purpose and Context of Each Evaluation	187

Chapter 12 •	Quantitative Evaluation Methods	218
Chapter 13 •	Qualitative Evaluation Methods	243
Chapter 14 •	Mixed-Method Evaluation	262
Chapter 15 •	Sampling Strategies for RealWorld Evaluation	289
Chapter 16 •	Evaluating Complex Projects, Programs, and Policies	315
Chapter 17 •	Gender Evaluation: Integrating Gender Analysis Into Evaluations	353
Chapter 18 •	Evaluation in the Age of Big Data	382
 PART III • MANAGING EVALUATIONS		
Chapter 19 •	Managing Evaluations	430
Chapter 20 •	The Road Ahead	459
 Glossary of Terms and Acronyms		474
References		487
Author Index		504
Subject Index		510

Appendices available online at
and also in the various eBook versions of the text.

DETAILED CONTENTS

List of Boxes, Figures, and Tables	xxiv
List of Appendices	xxxi
Foreword by Jim Rugh	xxxiii
Preface	xxxv
Acknowledgments	xlii
About the Authors	xliv

PART I • THE SEVEN STEPS OF THE REALWORLD EVALUATION APPROACH

Chapter 1 • Overview: RealWorld Evaluation and the Contexts in Which It Is Used	2
1. Welcome to RealWorld Evaluation	2
2. The RealWorld Evaluation Context	8
3. The Four Types of Constraints Addressed by the RealWorld Approach	9
3.1 Budget and Other Resource Constraints	9
3.2 Time Constraints	11
3.3 Data Constraints	11
3.4 Political and Organizational Influences and Constraints	12
4. Additional Organizational and Administrative Challenges	13
5. The RealWorld Approach to Evaluation Challenges	14
6. Who Uses RealWorld Evaluation, for What Purposes, and When?	16
Summary	18
Further Reading	18
Chapter 2 • First Clarify the Purpose: Scoping the Evaluation	19
1. Stakeholder Expectations of Impact Evaluations	20
2. Understanding Information Needs	22
3. Developing the Program Theory Model	25
3.1 Theory-Based Evaluation (TBE) as a Management Tool	30
4. Identifying the Constraints to Be Addressed by RWE and Determining the Appropriate Evaluation Design	30
5. Developing Designs Suitable for RealWorld Evaluation Conditions	31
5.1 How the Availability of Data Affects the Choice of Evaluation Design	32
5.2 Developing the Terms of Reference (Statement of Work) for the Evaluation	36
Summary	37
Further Reading	37

Chapter 3 • Not Enough Money: Addressing Budget Constraints	39
1. Simplifying the Evaluation Design	39
1.1 Simplifying the Design for Quantitative Evaluations	42
1.2 Simplifying the Design for Qualitative Evaluations	44
2. Clarifying Client Information Needs	46
3. Using Existing Data	47
4. Reducing Costs by Reducing Sample Size	47
4.1 Adjusting the Sample Size Based on Client Information Needs and the Kinds of Decisions to Which the Evaluation Will Contribute	47
4.2 Factors Affecting Sample Size for Quantitative Evaluations	48
<i>Effect of the Level of Disaggregation on the Required Sample Size</i>	49
4.3 Factors Affecting the Size of Qualitative Samples	50
4.4 Factors Affecting the Size of Mixed-Method Samples	50
5. Reducing Costs of Data Collection and Analysis	50
6. Assessing the Feasibility and Utility of Using New Information Technology (NIT) to Reduce the Costs of Data Collection	52
7. Threats to Validity of Budget Constraints	53
Summary	53
Further Reading	54
 Chapter 4 • Not Enough Time: Addressing Scheduling and Other Time Constraints	 55
1. Similarities and Differences Between Time and Budget Constraints	55
2. Simplifying the Evaluation Design	59
3. Clarifying Client Information Needs and Deadlines	60
4. Using Existing Documentary Data	61
5. Reducing Sample Size	61
6. Rapid Data-Collection Methods	61
7. Reducing Time Pressure on Outside Consultants	67
8. Hiring More Resource People	68
9. Building Outcome Indicators Into Project Records	68
10. New Information Technology for Data Collection and Analysis	69
11. Common Threats to Adequacy and Validity Relating to Time Constraints	70
Summary	70
Further Reading	71
 Chapter 5 • Critical Information Is Missing or Difficult to Collect: Addressing Data Constraints	 72
1. Data Issues Facing RealWorld Evaluators	72
2. Reconstructing Baseline Data	76
2.1 Strategies for Reconstructing Baseline Data	76
<i>When Should Baseline Data Be Collected?</i>	76
<i>Using Administrative Data From the Project</i>	77
<i>Cautionary Tales—and Healthy Skepticism</i>	79
<i>Using Other Sources of Secondary Data</i>	80

<i>Conducting Retrospective Surveys</i>	81
<i>Working With Key Informants</i>	83
<i>Using Participatory Evaluation Methods</i>	83
<i>Using Geographical Information Systems (GIS) and Satellite Images to Reconstruct Baseline Data</i>	84
3. Special Issues Reconstructing Baseline Data for Project Populations and Comparison Groups	85
3.1 Challenges Collecting Baseline Data on the Project Group	85
3.2 Challenges Collecting Baseline Data on a Comparison Group	86
3.3 The Challenge of Omitted Variables (“Unobservables”)	86
4. Collecting Data on Sensitive Topics or From Difficult-to-Reach Groups	88
4.1 Addressing Sensitive Topics	88
4.2 Studying Difficult-to-Reach Groups	88
5. Common Threats to Adequacy and Validity of an Evaluation Relating to Data Constraints	92
Summary	92
Further Reading	93
Chapter 6 • Political Constraints	94
1. Values, Ethics, and Politics	94
2. Societal Politics and Evaluation	95
3. Stakeholder Politics	97
4. Professional Politics	98
4.1 Teaming	98
5. Political Issues in the Design Phase	99
5.1 Hidden Agendas and Pseudo-Evaluation	99
5.2 Stakeholder Differences	100
6. Political Issues in the Conduct of an Evaluation	100
6.1 Shifting Evaluator Roles	100
6.2 Data Access	101
7. Political Issues in Evaluation Reporting and Use	101
7.1 Bias in Reporting	102
7.2 Use and Misuse of Findings	102
8. Advocacy	102
8.1 Activism	103
8.2 Strategizing for Use	103
Summary	104
Further Reading	104
Chapter 7 • Strengthening the Evaluation Design and the Validity of the Conclusions	105
1. Validity in Evaluation	105
2. Factors Affecting Adequacy and Validity	107
3. A Framework for Assessing the Validity and Adequacy of QUANT, QUAL, and Mixed-Method Designs	108
3.1 The Categories of Validity (Adequacy, Trustworthiness)	109

4. Assessing and Addressing Threats to Validity for Quantitative Impact Evaluations	112
4.1 A Threats-to-Validity Worksheet for QUANT Evaluations	112
4.2 Strengthening Validity in Quantitative Evaluations by Strengthening the Evaluation Design	114
<i>Random Sampling</i>	114
<i>Triangulation</i>	115
<i>Selection of Statistical Procedures</i>	116
<i>Peer Review and Meta-Evaluation</i>	116
4.3 Taking Corrective Actions When Threats to Validity Have Been Identified	116
5. Assessing Adequacy and Validity for Qualitative Impact Evaluations	118
5.1 Strengthening Validity in Qualitative Evaluations by Strengthening the Evaluation Design	120
<i>Purposeful (Purposive) Sampling</i>	120
<i>Triangulation</i>	120
<i>Validation</i>	121
<i>Meta-Evaluation and Peer Review</i>	121
5.2 Addressing Threats to Validity in Qualitative Evaluations	121
<i>Collecting Data Across the Full Range of Appropriate Settings, Times, and Respondents</i>	121
<i>Inappropriate Subject Selection</i>	122
<i>Insufficient Language or Cultural Skills to Ensure Sensitivity to Informants</i>	122
<i>Insufficient Opportunity for Ongoing Analysis by the Team</i>	122
<i>Minimizing Observer Effects</i>	122
6. Assessing Validity for Mixed-Method (MM) Evaluations	122
7. Using the Threats-to-Validity Worksheets	123
7.1 Points During the RWE Cycle at Which Corrective Measures Can Be Taken	124
<i>Strengthening the Evaluation Design</i>	125
Summary	125
Further Reading	126

Chapter 8 • Making It Useful: Helping Clients and Other Stakeholders Utilize the Evaluation **127**

1. What Do We Mean by Influential Evaluations and Useful Evaluations?	127
2. The Underutilization of Evaluation Studies	130
2.1 Why Are Evaluation Findings Underutilized?	131
2.2 The Challenges of Utilization for RWE	131
3. Strategies for Promoting the Utilization of Evaluation Findings and Recommendations	132
3.1 The Importance of the Scoping Phase	133
<i>Understand the Client's Information Needs</i>	133
<i>Understand the Dynamics of the Decision-Making Process and the Timing of the Different Steps</i>	135
<i>Define the Program Theory on Which the Program Is Based in Close Collaboration With Key Stakeholders</i>	135
<i>Identify Budget, Time, and Data Constraints and Prioritize Their Importance and the Client's Flexibility to Adjust Budget or Time If Required to Improve the Quality of the Evaluation</i>	136

<i>Understand the Political Context</i>	136
<i>Prepare a Set of RWE Design Options to Address the Constraints and to Strategize</i>	
<i>With the Client to Assess Which Option Is Most Acceptable</i>	136
3.2 Formative Evaluation Strategies	136
3.3 Communication With Clients Throughout the Evaluation	138
3.4 Evaluation Capacity Building	138
3.5 Strategies for Overcoming Political and Bureaucratic Challenges	138
3.6 Communicating Findings	139
3.7 Developing a Follow-Up Action Plan	141
Summary	141
Further Reading	141

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	144
1. Standards of Competence	144
2. Professional Standards	144
2.1 Evaluation Standards	145
2.2 Guiding Principles	146
3. Ethical Codes of Conduct	146
3.1 International Standards of Ethics	147
3.2 National Standards of Ethics	147
3.3 Ethical Frameworks in Evaluation	148
4. Issues	149
4.1 Advocacy	149
4.2 Accountability	149
4.3 International Regulation	150
Summary	151
Further Reading	151
Chapter 10 • Theory-Based Evaluation and Theory of Change	153
1. Theory-Based Evaluation (TBE) and Theory of Change (TOC)	153
1.1 Program Theory and Theory-Based Evaluation	153
<i>Stage 1: Articulation of the Program Theory Model</i>	154
<i>Stage 2: The Results Framework</i>	154
<i>Stage 3: The Logical Framework</i>	154
<i>Stages 4a and 4b: Impact and Implementation Models</i>	156
<i>Applications of Program Theory</i>	156
1.2 Theory of Change	156
<i>Benefits of the TOC for Program Evaluation</i>	158
2. Applications of Program Theory in Program Evaluation	160
2.1 The Increasing Use of Program Theory in Evaluation	160
2.2 Examples of TBE Tools and Techniques That Can Be Applied in Policy, Program, and Project Evaluations and at Different Stages of the Program Cycle	162
2.3 Realist Evaluation	164

3. Using TOC in Program Evaluation	165
3.1 Conceptualizing the Change Process	165
3.2 Identifying and Testing Hypotheses About the Processes of Implementation and Change	165
3.3 Identifying Unintended Outcomes	165
3.4 Addressing Complexity and Emergence	166
<i>Addressing Emergence</i>	167
3.5 Challenges Affecting the Use of TOCs	167
4. Designing a Theory of Change Evaluation Framework	168
4.1 The Different Ways That a Theory of Change Can Be Used	168
4.2 The Purpose of the Theory of Change	169
4.3 Representing the Theory of Change	170
<i>An Example of a Theory of Change</i>	172
4.4 Designing the TOC and the Sources of Data	172
5. Integrating a Theory of Change Into the Program Management, Monitoring, and Evaluation Cycle	175
5.1 Articulating the Program Theory	175
5.2 Process Tracing	177
5.3 Results-Based Reporting and Logical Frameworks	177
5.4 Program Impact and Implementation Models	179
6. Program Theory Evaluation and Causality	179
6.1 The Debate on the Value of Program Theory to Explain Causality	179
6.2 Using Mixed-Method Program Theory Designs to Explain Causality	181
Summary	184
Further Reading	184

Chapter 11 • Evaluation Designs: The RWE Strategy for Selecting the Appropriate Evaluation Design to Respond to the Purpose and Context of Each Evaluation **187**

1. Different Approaches to the Classification of Evaluation Designs	187
2. Assessing Causality Attribution and Contribution	190
2.1 Attribution Analysis and Contribution Analysis	190
2.2 An Introduction to Contribution Analysis and Outcome Harvesting	190
3. The RWE Approach to the Selection of the Appropriate Impact Evaluation Design	192
3.1 Design Step 1	193
<i>Using the Evaluation Purpose and Context Checklist</i>	193
3.2 Design Step 2	197
<i>Analysis of the Evaluation Design Framework</i>	197
3.3 Design Step 3	201
<i>Identify a Short List of Potential Evaluation Designs</i>	201
3.4 Design Step 4	204
<i>Take Into Consideration the Preferred Methodological Approaches of Stakeholders and Evaluators</i>	204
3.5 Design Step 5	205
<i>Strategies for Strengthening the Basic Evaluation Designs</i>	205
3.6 Design Step 6	205
<i>Evaluability Analysis to Assess the Technical, Resource, and Political Feasibility of Each Design</i>	205

3.7 Design Step 7	205
<i>Preparation of Short List of Evaluation Design Options for Discussion With Clients and Other Stakeholders</i>	205
3.8 Design Step 8	207
<i>Agreement on the Final Evaluation Design</i>	207
4. Tools and Techniques for Strengthening the Basic Evaluation Designs	208
4.1 Basing the Evaluation on a Theory of Change and a Program Theory Model	208
4.2 Process Analysis	209
4.3 Incorporating Contextual Analysis	209
4.4 Complementing Quantitative Data Collection and Analysis With Mixed-Method Designs	210
4.5 Making Full Use of Available Secondary Data	210
4.6 Triangulation: Using Two or More Independent Estimates for Key Indicators and Using Data Sources and Analytical Methods to Explain Findings	210
5. Selecting the Best Design for RealWorld Evaluation Scenarios	210
5.1 Factors Affecting the Choice of the Appropriate Design for a Particular Evaluation	210
5.2 Challenges Facing the Use of Experimental and Other Statistical Designs in RealWorld Evaluation Contexts	213
5.3 Selecting the Appropriate Designs for RealWorld Evaluation Scenarios	213
Summary	215
Further Reading	216
 Chapter 12 • Quantitative Evaluation Methods	 218
1. Quantitative Evaluation Methodologies	218
1.1 The Importance of Program Theory in the Design and Analysis of QUANT Evaluations	218
1.2 Quantitative Sampling	219
2. Experimental and Quasi-Experimental Designs	219
2.1 Randomized Control Trials (RCTs)	219
2.2 Quasi-Experimental Designs (QEDs)	223
3. Strengths and Weaknesses of Quantitative Evaluation Methodologies	226
4. Applications of Quantitative Methodologies in Program Evaluation	226
4.1 Analysis of Population Characteristics	226
4.2 Hypothesis Testing and the Analysis of Causality	226
4.3 Cost-Benefit Analysis and the Economic Rate of Return (ERR)	228
4.4 Cost-Effectiveness Analysis	228
5. Quantitative Methods for Data Collection	229
5.1 Questionnaires	229
<i>Types of Questions Used in Quantitative Surveys</i>	230
5.2 Interviewing	230
5.3 Observation	230
<i>Observational Protocols</i>	230
<i>Unobtrusive Measures in Observation</i>	231
5.4 Focus Groups	231
5.5 Self-Reporting Methods	232
5.6 Knowledge and Achievement Tests	232

5.7 Anthropometric and Other Physiological Health Status Measures	232
5.8 Using Secondary Data	233
<i>Common Problems With Secondary Data for Evaluation Purposes</i>	234
6. The Management of Data Collection for Quantitative Studies	235
6.1 Survey Planning and Design	235
6.2 Implementation and Management of Data Collection	236
<i>RealWorld Constraints on the Management of Data Collection</i>	237
7. Data Analysis	239
Summary	240
Further Reading	240
 Chapter 13 • Qualitative Evaluation Methods	 243
1. Design	243
1.1 Flexibility	244
1.2 Contextuality	245
1.3 Sensitivity	245
2. Data Collection	245
2.1 Observation	246
<i>Collecting Observation Data</i>	246
<i>Structure in Observations</i>	246
<i>Participant Observations</i>	247
<i>Writing Up Observation Data</i>	250
2.2 Interview	250
<i>Collecting Interview Data</i>	250
<i>Structure in Interviews</i>	251
<i>Group Interviews</i>	251
<i>Hybrid Interview Methods</i>	252
<i>Writing Up Interview Data</i>	253
2.3 Analysis of Documents and Artifacts	253
2.4 Technology in Data Collection	254
2.5 Triangulation and Validation	254
<i>Triangulation</i>	255
<i>Validation</i>	255
3. Data Analysis	256
3.1 Inductive Analysis	256
<i>Constant-Comparative Method</i>	256
3.2 Thematic Analysis	256
<i>Electronic Data Analysis</i>	256
3.3 Holistic Analysis	257
<i>Data Complexity</i>	257
<i>Criteriality</i>	257
3.4 Validity	257
<i>Disciplining Subjective Judgment</i>	258
<i>Theoretical Triangulation</i>	258
<i>Peer Review and Meta-Evaluation</i>	258
3.5 Generalizability	259
4. Reporting	259
5. Real-World Constraints	260

Summary	260
Further Reading	261
Chapter 14 • Mixed-Method Evaluation	262
1. The Mixed-Method Approach	262
2. Rationale for Mixed-Method Approaches	263
2.1 Why Use Mixed Methods?	263
2.2 Areas Where Mixed Methods Can Potentially Strengthen Evaluations	267
<i>Improving Construct Validity and Data Quality</i>	267
<i>Evaluating Complex Programs</i>	268
<i>Strengthening Big Data–Based Evaluations</i>	268
<i>Identifying Unintended Outcomes (UOs) of Development Programs</i>	269
3. Approaches to the Use of Mixed Methods	269
3.1 Applying Mixed Methods When the Dominant Design Is Quantitative or Qualitative	273
3.2 Using Mixed Methods When Working Under Budget, Time, and Data Constraints	275
4. Mixed-Method Strategies	276
4.1 Sequential Mixed-Method Designs	276
4.2 Concurrent Designs	278
<i>Concurrent Triangulation Design</i>	279
<i>Concurrent Nested Design</i>	280
4.3 Using Mixed Methods at Different Stages of the Evaluation	281
5. Implementing a Mixed-Method Design	281
<i>Composition of the Research Team</i>	283
<i>Using Integrated Approaches at Different Stages of the Evaluation</i>	283
6. Using Mixed Methods to Tell a More Compelling Story of What a Program Has Achieved	284
7. Case Studies Illustrating the Use of Mixed Methods	286
Summary	286
Further Reading	287
Chapter 15 • Sampling Strategies for RealWorld Evaluation	289
1. The Importance of Sampling for RealWorld Evaluation	290
2. Purposive Sampling	291
2.1 Purposive Sampling Strategies	292
2.2 Purposive Sampling for Different Types of Qualitative Data Collection and Use	294
<i>Sampling for Data Collection</i>	294
<i>Sampling Data for Qualitative Reporting</i>	295
2.3 Considerations in Planning Purposive Sampling	295
3. Probability (Random) Sampling	296
3.1 Key Questions in Designing a Random Sample for Program Evaluation	296
3.2 Selection Procedures in Probability (Random) Sampling	299
3.3 Sample Design Decisions at Different Stages of the Survey	300
<i>Presampling Questions</i>	300
<i>Questions and Choices During the Sample Design Process</i>	301
<i>Postsampling Questions and Choices</i>	302

3.4 Sources of Error in Probabilistic Sample Design	303
<i>Nonsampling Bias</i>	303
<i>Sampling Bias</i>	304
4. Using Power Analysis and Effect Size for Estimating the Appropriate Sample Size for an Impact Evaluation	305
4.1 The Importance of Power Analysis for Determining Sample Size for Probability Sampling	305
4.2 Defining Effect Size	306
4.3 Type I and Type II Errors	306
4.4 Defining the Power of the Test	307
4.5 An Example Illustrating the Relationship Between the Power of the Test, the Effect Size, and the Required Sample Size	307
5. The Contribution of Meta-Analysis	309
6. Sampling Issues for Mixed-Method Evaluations	309
6.1 Model 1: Using Mixed Methods to Strengthen a Mainly Quantitative Evaluation Design	309
6.2 Model 2: Using a Mixed-Method Design to Strengthen a Qualitative Evaluation Design	311
6.3 Model 3: Using an Integrated Mixed-Method Design	311
7. Sampling Issues for RealWorld Evaluation	312
7.1 Lot Quality Acceptance Sampling (LQAS): An Example of a Sampling Strategy Designed to Be Economical and Simple to Administer and Interpret	312
Summary	313
Further Reading	314

Chapter 16 • Evaluating Complex Projects, Programs, and Policies **315**

1. The Move Toward Complex, Country-Level Development Programming	315
2. Defining Complexity in Development Programs and Evaluations	317
2.1 The Dimensions of Complexity	317
<i>Dimension 1: The Nature of the Intervention</i>	317
<i>Dimension 2: Institutions and Stakeholders</i>	319
<i>Dimension 3: The Context (System) Within Which the Program Is Implemented</i>	319
<i>Dimension 4: Causality and Change</i>	320
<i>The Complexity of the Evaluation</i>	320
2.2 Simple Projects, Complicated Programs, and Complex Development Interventions	321
2.3 The Main Types of Complex Interventions	326
2.4 Assessing Levels of Complexity	327
<i>How to Use the Checklist</i>	330
2.5 Special Challenges for the Evaluation of Complex, Country-Level Programs	330
3. A Framework for the Evaluation of Complex Development Programs	331
3.1 Overview of the Complexity-Responsive Evaluation Framework	331
3.2 Drawing on Big Data Science	333
3.3 Step 1: Mapping the Dimensions of Complexity	333
3.4 Step 2: Choosing a Unit of Analysis for Unpacking Complex Interventions Into Evaluable Components	338

Option 1: Implementation Components, Phases, and Themes	338
Option 2: Program Theories	339
Option 3: Cases	339
Option 4: Variables	340
3.5 Step 3: Choosing an Evaluation Design for Evaluating Each Unpacked Component	340
Attribution and the Challenge of Defining the Counterfactual for Complex Evaluations	340
Choosing the Best Evaluation Design for Evaluating the Unpacked Program Components	340
Operational Unpacking of Components of Complex Interventions	342
Program Theory and Theory Reconstruction	343
Case Studies and Rich Description	344
Variable-Based Approaches to Unpacking	344
3.6 Step 4: Choosing an Approach to Reassembling the Various Parts Into a Big Picture	345
Systems Modeling	345
Descriptive and Inferential Statistical Analysis	346
Comparative Case Study Approaches	346
Portfolio Analysis	347
Review and Synthesis Approaches	347
Rating Scales	348
3.7 Step 5: Assessing Program Contribution to National and International Development Goals—Going Back to the Big Picture	350
Summary	351
Further Reading	352

Chapter 17 • Gender Evaluation: Integrating Gender Analysis Into Evaluations **353**

1. Why a Gender Focus Is Critical	354
1.1 Why So Few Evaluations Have a Gender Focus	355
Political Constraints	355
Skill and Resource Constraints	355
Methodological Constraints	356
1.2 The Value of a Gender Evaluation	356
2. Gender Issues in Evaluations	357
Gender Does Not Just Focus on Women	357
Intersectionality	357
Multisectoral Analysis	357
Complexity and a Longitudinal Perspective	358
Gender Approaches Are Normative	358
Bodily Integrity and Sexuality	359
Systems of Social Control and Gender	359
Defining Boundaries for the Evaluation	360
2.1 How to “Gender” an Evaluation	361
Using a Gender Analysis Framework	361
Using a Feminist Evaluation Approach	363
2.2 A Continuum of Gendered Evaluation	363
Level 1: Sex Disaggregation of a Set of Basic Indicators	364
Level 2: Analysis of Factors Affecting Women’s Participation in Development	365
Level 3: Analysis of Household, Community, and Social Dynamics	365
Level 4: Comprehensive Feminist Analysis for Empowerment	366

3. Designing a Gender Evaluation	366
3.1 Design Criteria	366
<i>Evaluation Criteria—Adapting the OECD/DAC Framework</i>	366
3.2 Selecting Projects for a Gendered Evaluation	367
<i>Institutional Intervention Points</i>	368
<i>Portfolio Analysis and Meta-Analysis</i>	368
<i>Gender Flags and Checklists</i>	368
3.3 Defining the Depth and Scope	369
<i>Depth of the Analysis</i>	369
<i>Boundaries of the Evaluation</i>	369
<i>Time Horizons</i>	370
3.4 Defining Evaluation Questions	371
<i>The Importance of Broad-Based Stakeholder Consultations</i>	372
3.5 Evaluability Assessment	372
4. Gender Evaluations With Different Scopes	373
4.1 Complex Development Programs Evaluation	373
4.2 Single-Country Evaluations	373
4.3 Multicountry Evaluations	374
4.4 Country Gender M&E Framework	374
4.5 Gender in Sustainable Development Goals	374
5. The Tools of Gender Evaluation	376
5.1 Advanced Designs for Gender Evaluations	377
5.2 Attribution and Contribution Analysis	378
Summary	378
Further Reading	379

Chapter 18 • Evaluation in the Age of Big Data **382**

1. Introducing Big Data and Data Science	383
1.1 Increasing Application of Big Data in Personal and Public Life	383
1.2 Defining Big Data, Data Science, and New Information Technology (NIT)	384
<i>Big Data Is Embedded in New Information Technology</i>	384
<i>Defining Big Data</i>	387
1.3 The Data Continuum	392
1.4 The Position of Evaluation in the Big Data Ecosystem	393
<i>The Big Data Ecosystem</i>	393
<i>The Evaluation Ecosystem</i>	394
2. Increasing Application of Big Data in the Development Context	396
3. The Tools of Data Science	398
3.1 The Stages of the Data Analytics Cycle	398
<i>Step 1: Descriptive and Exploratory Analysis: Documenting What Is Happening, Often in Real Time</i>	399
<i>Step 2: Predictive Analysis: What Is Likely to Happen</i>	401
<i>Step 3: Detection: Tracking Who Is Likely to Succeed and Who Will Fail</i>	403
<i>Step 4: Prescription: Evaluating How Outcomes Were Achieved and Providing Recommendations on How to Improve Program Performance</i>	403
3.2 The Tools of Data Analytics	404
3.3 The Limitations of Data Science	408
<i>Being Aware of, and Addressing, Methodological Challenges</i>	408
<i>Political and Organizational Challenges</i>	408

<i>Privacy and Security</i>	409
<i>New Ethical Challenges</i>	409
<i>The Dark Side of Big Data</i>	410
4. Potential Applications of Data Science in Development Evaluation	411
4.1 Challenges Facing Current Evaluation Approaches	411
<i>Data Analysis Challenges</i>	412
4.2 How Data Science Can Strengthen Evaluation Practice	413
5. Building Bridges Between Data Science and Evaluation	417
5.1 The Potential Benefits of Convergence of Data Science and Evaluation	417
5.2 The Need for Caution When Assessing the Benefits of Big Data and Data Analytics for Development Evaluation	419
5.3 Challenges	420
5.4 Skills Required for Present and Future Evaluators	422
<i>Understanding the Big Debates Around the Evaluation of Future Development Programs</i>	422
<i>New Skills Required for Evaluation Offices and Evaluators and for Data Scientists</i>	423
<i>Skills Development for Evaluators</i>	423
<i>Evaluation Skills for Data Scientists</i>	424
5.5 Necessary Conditions for Convergence to Occur	424
<i>To What Extent Do These Conditions Exist in Both Industrial and Developing Countries?</i>	425
5.6 Possible Collaborative Bridge-Building Initiatives	425
Summary	425
Further Reading	426

PART III • MANAGING EVALUATIONS

Chapter 19 • Managing Evaluations	430
1. Organizational and Political Issues Affecting the Design, Implementation, and Use of Evaluations	430
2. Planning and Managing the Evaluation	431
2.1 Step 1: Preparing the Evaluation	431
<i>Step 1-A: Defining the Evaluation Framework or the Scope of Work (SoW)</i>	431
<i>Step 1-B: Linking the Evaluation to the Project Results Framework</i>	433
<i>Step 1-C: Involving Stakeholders</i>	433
<i>Step 1-D: Commissioning Diagnostic Studies</i>	433
<i>Step 1-E: Defining the Management Structure for the Evaluation</i>	435
2.2 Step 2: Recruiting the Evaluators	436
<i>Step 2-A: Recruiting the Internal Evaluation Team</i>	436
<i>Step 2-B: Different Ways to Contract External Evaluation Consultants</i>	436
<i>Step 2-C: Preparing the Request for Proposals (RFP)</i>	437
<i>Step 2-D: Preparing the Terms of Reference (ToR)</i>	438
<i>Step 2-E: Selecting the Consultants</i>	439
2.3 Step 3: Designing the Evaluation	439
<i>Step 3-A: Formulating Evaluation Questions</i>	439
<i>Step 3-B: Assessing the Evaluation Scenario</i>	441
<i>Step 3-C: Selecting the Appropriate Evaluation Design</i>	441
<i>Step 3-D: Commissioning an Evaluability Assessment</i>	441
<i>Step 3-E: Designing "Evaluation-Ready" Programs</i>	442

2.4 Step 4: Implementing the Evaluation	443
<i>Step 4-A: The Role of the Evaluation Department of a Funding Agency, Central Government Ministry, or Sector Agency</i>	443
<i>Step 4-B: The Inception Report</i>	444
<i>Step 4-C: Managing the Evaluation</i>	444
<i>Step 4-D: Working With Stakeholders</i>	445
<i>Step 4-E: Quality Assurance (QA)</i>	446
<i>Step 4-F: Management Challenges With Different Kinds of Evaluation</i>	447
2.5 Step 5: Reporting and Dissemination	449
<i>Step 5-A: Providing Feedback on the Draft Report</i>	449
<i>Step 5-B: Disseminating the Evaluation Report</i>	450
2.6 Step 6: Ensuring Implementation of the Recommendations	450
<i>Step 6-A: Coordinating the Management Response and Follow-Up</i>	450
<i>Step 6-B: Facilitating Dialogue With Partners</i>	451
3. Institutionalizing Impact Evaluation Systems at the Country and Sector Levels	451
3.1 Institutionalizing Impact Evaluation	451
3.2 Integrating IE Into Sector and/or National M&E and Other Data-Collection Systems	453
<i>Creating Demand for IE</i>	454
4. Evaluating Capacity Development	454
Summary	455
Further Reading	457
Chapter 20 • The Road Ahead	459
1. Conclusions	459
1.1 The Challenge of Assessing Impacts in a World in Which Many Evaluations Have a Short-Term Focus	459
1.2 The Continuing Debate on the “Best” Evaluation Methodologies	459
1.3 Selecting the Appropriate Evaluation Design	461
1.4 Mixed Methods: The Approach of Choice for Most RealWorld Evaluations	463
1.5 How Does RealWorld Evaluation Fit Into the Picture?	463
1.6 Quality Assurance	463
1.7 Need for a Strong Focus on Gender Equality and Social Equity	464
1.8 Basing the Evaluation Design on a Program Theory Model	465
1.9 The Importance of Context	465
1.10 The Importance of Process	466
1.11 Dealing With Complexity in Development Evaluation	466
1.12 Emergence	467
1.13 Integrating the New Information Technologies Into Evaluation	468
1.14 Greater Attention Must Be Given to the Management of Evaluations	468
1.15 The Challenge of Institutionalization of Evaluation	469
1.16 The Importance of Competent Professional and Ethical Practice	470
2. Recommendations	470
2.1 Developing Standardized Methodologies for the Evaluation of Complex Programs	470
2.2 Creative Approaches for the Definition and Use of Counterfactuals	471
2.3 Strengthening Quality Assurance and Threats to Validity Analysis	471

2.4 Defining Minimum Acceptable Quality Standards for Conducting Evaluations Under Constraints	471
2.5 Further Refinements to Program Theory	472
2.6 Further Refinements to Mixed-Method Designs	472
2.7 Integrating Big Data and Data Science Into Program Evaluation	472
2.8 Further Work Is Required to Strengthen the Integration of a Gender-Responsive Approach Into Evaluation Programs	473
2.9 The Road Ahead	473
Glossary of Terms and Acronyms	474
References	487
Author Index	504
Subject Index	510

Appendices available online at

and also in the various eBook versions of the text.

LIST OF BOXES, FIGURES, AND TABLES

BOXES

Box 1.1	The Implications of New Information Technology for RealWorld Evaluation	5
Box 1.2	A Familiar Evaluation Story	9
Box 1.3	Budget and Time Constraints Have Different Implications for the Evaluation Design	11
Box 1.4	Problems in Capturing Information From or About Women	12
Box 1.5	Political Influence on the Evaluation of a Hydroelectric Program in Asia	13
Box 1.6	How Organizational and Administrative Considerations Can Affect Budget, Time, Data, and Political Constraints	14
Box 1.7	Getting Inside the “Black Box”	15
Box 2.1	Boundaries and Boundary Choices	20
Box 2.2	Defining Outcomes and Impacts	29
Box 2.3	The Main Kinds of Evaluation Design	31
Box 3.1	Six Questions to Help Make the Budget Go Farther	42
Box 3.2	“Robust” Designs Require Correct Implementation!	44
Box 3.3	Examples of Economical Methods of Data Collection	52
Box 4.1	Two Kinds of Time Saving	59
Box 4.2	Time Constraints Affecting the Evaluation of a Water and Sanitation Project in Mozambique	59
Box 4.3	Examples of Rapid Data-Collection Methods	65
Box 4.4	Rapid Ethnographic Methods (RAP)	66
Box 4.5	Reducing the Time Required to Conduct a Mixed-Method Study of Survival Strategies of Low-Income Households	66
Box 5.1	Examples of Sampling Frames That May Exclude Sectors of the Intended Target Population	75
Box 5.2	Using Administrative Data From the Project to Reconstruct the Baseline	78
Box 5.3	Examples From Developing Countries of Databases That Do Not Adequately Represent Both Sexes	79
Box 5.4	Ghost Schools and Missing Girls: Why the Validity of Secondary Data Must Always Be Checked	80
Box 5.5	Examples of Retrospective Surveys Used to Reconstruct Baseline Data	82
Box 5.6	Selecting a Wide Range of Informants When Studying Sensitive Topics: Studying Crack Cocaine Addicts in New York	83
Box 5.7	Constructing a PRA Seasonal Calendar to Reconstruct Periods of Stress in a Rural Community in Kenya	85
Box 5.8	Potentially Strong and Weak Comparison Groups	87
Box 5.9	Examples of Collecting Data on Sensitive Topics	89

Box 6.1	Example of Difficulty in Obtaining Stakeholder Input	97
Box 6.2	Observing Staged Events and Exemplary Sites	101
Box 6.3	Spinning the Findings: An Old Art Form	102
Box 7.1	Factors Affecting the Validity of CARE International's Program Evaluation Methodology	107
Box 7.2	Ways to Address Typical Threats to Statistical, Internal Design, and Construct and External Validity for a Mixed-Method Evaluation	117
Box 7.3	Frick's Process-Oriented Approach to Assessing the Quality of Qualitative Evaluations	120
Box 8.1	How Are Evaluations Used? When Are They Influential?	129
Box 8.2	Political and Bureaucratic Challenges Affecting Use of Evaluations	132
Box 8.3	Nine Actions to Promote Evaluation Utilization	133
Box 8.4	Guidelines for Helping Clients Use the Evaluation	134
Box 8.5	Enhancing Evaluation Utilization—Lessons Learned	135
Box 8.6	Using Feedback From Process Analysis to Improve Project Performance (Formative Evaluation or Utilization-Focused Evaluation)	137
Box 8.7	An Eight-Step Strategy for Communicating the Findings of Social Science Research to Policymakers	140
Box 10.1	Two Approaches to the Theory of Change	157
Box 10.2	Anticipated Benefits of Working With Theory of Change	160
Box 10.3	Behavioral Economics as a Tool for Modeling Mechanisms of Change in TOC	166
Box 10.4	Six Common Challenges Facing the Articulation of Theories of Change	171
Box 10.5	Resources for Designing a Theory of Change	174
Box 10.6	Nine Potential Types of Evidence for Inferring Causality	182
Box 11.1	The Types of Evaluation Design That Will Be Covered in This and Subsequent Chapters	189
Box 11.2	Using Contribution Analysis to Assess the Effects of UNESCO Prizes	192
Box 11.3	Examples of Methodologically Strong Nonexperimental Designs	204
Box 11.4	How Program Theory Can Strengthen an Evaluation Design	208
Box 11.5	Examples of Big Data Tools and Techniques That Might Be Available to Strengthen the RWE Design	212
Box 11.6	Building Evidence in Challenging Contexts	214
Box 12.1	The Benefits of Randomized Designs	222
Box 12.2	Methods for Randomly Assigning Subjects to the Project and Control Groups	222
Box 12.3	Challenges in Using Randomized Control Trials	223
Box 13.1	Participant Observation in a Peer Mediation Program	249
Box 13.2	Group Interviews About Women's Use of Public Transportation	252
Box 13.3	Examples of Requests in Comprehensive Validation	255
Box 14.1	Elements of an Integrated, Multidisciplinary Research Approach	282
Box 15.1	Small Samples May Overlook Important Real Effects	291
Box 15.2	Be Cautious About Advice From Statisticians Not Familiar With the Subject Area	298

Box 15.3	Longitudinal Panel Studies Are Great—But What If You Can’t Locate the Original Sampled Households Two Years Later?	302
Box 15.4	Examples of How Big Data Can Strengthen Sample Designs	304
Box 16.1	A Hypothetical Case Study to Illustrate the Different Stages of the Evaluation Framework	332
Box 16.2	Using Satellite and Remote Sensor Images to Strengthen Sample Design for Evaluating the Impacts on Forest Cover of a Forest Protection Program	343
Box 16.3	Unpacking and Reassembling Using Qualitative Comparative Analysis	346
Box 17.1	Estimated Economic Costs of Gender Inequality: Some Examples From Africa, the Middle East, and Asia	354
Box 17.2	Oxfam’s Seven Drivers of Women’s Economic Empowerment	360
Box 17.3	Elements of a Gender Analysis Framework	362
Box 17.4	Six Fundamental Beliefs of Feminist Approaches	363
Box 17.5	A Gender Adaptation of the OECD/DAC Evaluation Criteria	367
Box 17.6	Institutional Intervention Points	368
Box 17.7	Case Studies of Gender Outcomes	369
Box 17.8	Examples of Gender Checklists	371
Box 17.9	Gender Equality Integration Into SDGs	375
Box 17.10	Methods for Gender Analysis Evaluation	377
Box 17.11	Methods for Feminist Evaluation	377
Box 18.1	We Are All Leaving Traceable Digital Trails of Our Lives	384
Box 18.2	Examples of Big Data Sources That Are, or Could Be, Used for Evaluation	386
Box 18.3	Top Finalists of the World Bank Big Data Innovation Challenge	397
Box 18.4	Using an Integrated Data Platform to Help End Modern Slavery	398
Box 18.5	Potential Use of Predictive Analytics Using Data From Government Flagship Programs in Developing Countries	399
Box 18.6	Applications of Machine Learning to Evaluation	400
Box 18.7	An Example of Social Media Analytics: Rapid Assessment of Communicable Disease Control Programs in Indonesia Based on the Analysis of Tweets	404
Box 18.8	Examples of the Application of Data Science Tools in Evaluation	417
Box 18.9	Using Drones in the Evaluation of an Irrigation Project	418
Box 20.1	Programs Can Dramatically Change During Implementation: The Example of a School Feeding Program	467

FIGURES

Figure 1.1	The RealWorld Evaluation Approach	7
Figure 2.1	Step 1: Scoping the Evaluation	21
Figure 2.2	A Simple Program Theory Model	27
Figure 3.1	Step 2: Addressing Budget and Other Resource Constraints	40
Figure 4.1	Step 3: Addressing Time Constraints	56
Figure 5.1	Step 4: Addressing Data Constraints	73

Figure 5.2	When Should the Baseline Study Be Conducted for the Evaluation of a Slum-Upgrading Project?	77
Figure 6.1	Step 5: Addressing Political Influences	96
Figure 7.1	Step 6: Strengthening the Evaluation Design and the Validity of Conclusions	106
Figure 7.2	Dimensions of Threats to Validity Used to Assess Quantitative and Qualitative Evaluation Designs	110
Figure 7.3	Dimensions of Threats to Validity for the Assessment of Mixed-Method Evaluation Designs	113
Figure 7.4	Structure, Audience, and Responsibility for the Different Parts of the Threats-to-Validity Worksheet	124
Figure 8.1	Step 7: Helping Clients Use the Evaluation	128
Figure 10.1	Theory of Change as a Component of a Program Theory	155
Figure 10.2	The Theory of Change as the “Missing Middle” of the Logic Model	159
Figure 10.3	Weaknesses of the “Black Box” Approach to Evaluation	159
Figure 10.4	Using a Results Chain to Identify Potential Benefits and Unintended Negative Outcomes of a Women’s Microcredit Program	161
Figure 10.5	The Elements of the Realist Evaluation Approach	164
Figure 10.6	A Theory of Change for Tackling Violence Against Women and Girls	173
Figure 10.7	Example of Process Tracing: Hypothetical Example Tracing the Effectiveness of Mechanisms to Strengthen the Quality of Evaluation Design on the Quality, and Accessibility to Vulnerable Groups, of Social Policies and Programs	178
Figure 10.8	Different Scenarios for the Variation of Project Effects Over Time	183
Figure 11.1	Eight Steps for the Selection of the Appropriate Evaluation Design to Address Different Evaluation Contexts, Purposes, and Scenarios	188
Figure 13.1	<i>Observation Protocol: Are Ethnic Minorities Excluded From Program Benefits?</i>	247
Figure 13.2	<i>Observation Protocol: Are Low Ability-Tracked Students Inappropriately Prevented From Moving to High Tracks?</i>	247
Figure 13.3	<i>Observation Protocol: Does Provision of Laptop Computers to Students Improve Achievement, Personalization, or Rigor of Their Educations or Classroom Community?</i>	248
Figure 13.4	Guidelines for Developing Questions for Interview Protocols	252
Figure 13.5	Documents and Artifacts That May Be Available From Evaluation Sites	254
Figure 14.1	Using a National Household Sample Frame to Ensure the Representativity and Credibility of a QUANT Case Study Evaluation Design Assessing the Access of Women to Rural Health Centers	265
Figure 14.2	Using QUAL Ethnographic Village Studies to Help Interpret the Findings of a National QUANT Sample Survey Estimating Women’s Utilization Rates of Rural Health Centers	266
Figure 14.3	The QUANT–QUAL Research Continuum	270
Figure 14.4	Sequential Mixed-Method Design With a Dominant Quantitative Approach: Studying Interhousehold Transfers of Money and Goods in Cartagena, Colombia	277
Figure 14.5	Sequential Mixed-Method Design With a Dominant Qualitative Approach: Evaluating the Adoption of New Seed Varieties by Different Types of Rural Families	278

Figure 14.6	Concurrent Triangulation Design: Comparing and Integrating Different Data Sources on Poverty and Household Income	279
Figure 14.7	Collaborative Outcome Reporting Technique (CORT): Data Collection, Analysis, and Report Structure	285
Figure 15.1	A Multistage Sample for Selecting a Sample of Secondary School Students	300
Figure 16.1	The Dimensions of Complexity of Development Programs and Their Evaluation	318
Figure 16.2	Simple, Linear Causal Chain	320
Figure 16.3	Complex Causal Chain	321
Figure 16.4	Examples of Complex, Nonproportional Change	322
Figure 16.5	Simple Projects, Complicated Programs, and Complex Interventions	323
Figure 16.6	The Five-Step Unpacking-Reassembling Approach for Evaluating Complex Programs	334
Figure 16.7	Unpacking the Micro-Grant Component of a Multidonor Decentralized Transport Support Program	335
Figure 16.8	Systems Map for the Grants Component of the Ministry of Transport Project	348
Figure 16.9	Disaggregating National Programs Into National, Regional, and Local Levels	349
Figure 17.1	The Ladder of Gendered Evaluation	364
Figure 17.2	A Continuum of Gendered Evaluation	375
Figure 18.1	Elements of the New Information Technology System	385
Figure 18.2	The Data Continuum	392
Figure 18.3	The Big Data Ecosystem and the Linkages to the Evaluation Ecosystem	395
Figure 18.4	The Four Stages of the Data Analytics Cycle	400
Figure 19.1	Managing the Evaluation Process	432
Figure 19.2	Three Pathways for the Evolution of Institutionalized IE Systems	452

TABLES

Table 1.1	RealWorld Evaluation Scenarios: Conducting Impact Evaluations With Time, Budget, Data, and Political and Organizational Constraints	10
Table 1.2	Who Uses RWE, for What Purposes, and When?	17
Table 2.1	Evaluation Design Components to Strengthen All of the Basic Evaluation Designs	33
Table 2.2	Determining Possible Evaluation Designs	34
Table 3.1	Reducing Costs of Data Collection and Analysis for Quantitative and Qualitative Evaluations	41
Table 3.2	Overview of Most Common RWE Impact Evaluation Designs Classified by Their Statistical Strength	45
Table 3.3	Estimated Cost Savings for Less Robust RWE Designs Compared With Design 2.2	46

Table 3.4	Sample Size Required for Various Minimum Detectable Effects (Decrease in Household Health Expenditures), Power = 0.9. No Clustering	49
Table 4.1	Options for Reducing the Time Required for Data Collection and Analysis in Quantitative and Qualitative Evaluations	57
Table 4.2	Rapid Data-Collection Methods: The Effect on Duration of the Evaluation and the Level of Effort	63
Table 5.1	Summary of Strategies for Addressing Data Constraints	91
Table 6.1	Points at Which Political Influences May Affect an Evaluation	95
Table 7.1	Factors Determining the Adequacy of the Evaluation Design and of the Findings	108
Table 10.1	Examples of the Application of TBE at Different Stages of the Evaluation Cycle	163
Table 10.2	The Different Ways in Which a Theory of Change Can Be Used	168
Table 10.3	Variations in the Estimated Level of Effects Depending on the Year in Which the Posttest Study Is Conducted	182
Table 11.1	The Contribution Analysis Six-Step Approach	191
Table 11.2	Evaluation Purpose and Context	194
Table 11.3	Seven Basic Impact Evaluation Designs and Two More Advanced Options	199
Table 11.4	Tools and Techniques to Strengthen Any of the Basic Impact Evaluation Designs	206
Table 12.1	Strengths and Weaknesses of QUANT Evaluation Approaches	227
Table 12.2	Features of Well-Designed Quantitative Evaluations	236
Table 14.1	Strengths and Weaknesses of Qualitative Evaluation Approaches	264
Table 14.2	Examples of Evaluation Designs at Each Point on the QUANT-QUAL Continuum	271
Table 15.1	Approximate Required Sample Size for Comparing a Single Sample Mean With the Population Mean to Achieve Different Levels of Power for a Range of Effect Sizes (With a One-Tail Test and Significance Level = 0.05)	308
Table 15.2	Mean Effect Sizes for Different Interventions Addressing Aggressive Behavior in Schools	309
Table 15.3	Estimated Minimum Sample Sizes Required for Different Types of Qualitative Research Components of Mixed-Method Designs	311
Table 15.4	Lot Quality Acceptance Sampling: Decision Rules for Determining the Minimum Number of Units With Satisfactory Delivery of Information or Services to Accept That a Particular Benchmark Level of Satisfactory Services Has Been Achieved (β errors > 10%)	313
Table 16.1	Complicated and Complex Aspects of Interventions	326
Table 16.2	Checklist for Assessing the Level of Complexity of a Program	328
Table 16.3	The Potential Applications of Complexity Science in the Complexity-Responsive Evaluation Approach (Framework)	336
Table 16.4	Alternative Approaches for Defining the Counterfactual With Complex Program Evaluations	341

Table 17.1	Scenarios for the Integration of Gender Into Evaluation	367
Table 18.1	Defining Big Data	388
Table 18.2	Comparing Big Data and Conventional Evaluation Data	389
Table 18.3	Widely Used NIT and Data Analytics Applications in International Development	396
Table 18.4	Kinds of Big Data Analysis With Potential Applications for Program Monitoring and Evaluation	402
Table 18.5	Some Widely Cited Categories of Data Analytics (Data Science)	407
Table 18.6	Potential Ways That Big Data and ICTs Can Strengthen Program Evaluation	413
Table 18.7	Examples Illustrating How Big Data and ICTs Have Been Used to Strengthen Widely Used Evaluation Designs	416
Table 19.1	Roles of Stakeholders in an Evaluation	434
Table 19.2	How Lack of Coordination Between Project Management and the Evaluator Can Weaken the Evaluation Design	443

LIST OF APPENDICES

- 2.1** A Checklist to Assess the Evaluation Purpose and the Context Within Which It Will Be Implemented
- 2.2** Seven Basic Impact Evaluation Designs
- 2.3** Potential Methodological Weaknesses in Many Statistically Strong Evaluation Designs
- 2.4** Developing the Terms of Reference (Statement of Work) for the Evaluation
- 3.1** Factors Affecting the Sample Size
- 3.2** Threats to Adequacy and Validity Relating to Budget Constraints
- 5.1** Working With Comparison Groups in Retrospective Evaluations
- 5.2** Using Recall Techniques in Retrospective Surveys
- 5.3** Challenges Collecting Baseline Data on a Comparison Group
- 5.4** Special Issues and Challenges Working With Comparison Groups
- 7.1** Worksheet for Assessing Threats to the Validity of the Findings and Recommendations of Quantitative (Experimental and Quasi-Experimental) Impact Evaluation Designs
- 7.2** Worksheet for Assessing Threats to the Validity of the Findings and Recommendations of Qualitative Impact Evaluation Designs
- 7.3** Integrated Worksheet for Assessing Threats to the Validity of the Findings and Recommendations of Mixed-Method Impact Evaluation Designs (Standard Version)
- 7.4** Example of a Completed Threats-to-Validity Worksheet
- 7.5** Integrated Worksheet for Assessing Threats to the Validity of the Findings and Recommendations of Mixed-Method Impact Evaluation Designs (Advanced Version)
- 7.6** Approaches for Assessing Validity of Mixed-Method Evaluations
- 7.7** Points During the RWE Cycle at Which Corrective Measures Can Be Taken
- 7.8** Factors Determining the Adequacy of an Evaluation Design and the Validity of the Findings
- 7.9** Examples of Other Checklists Used to Assess Evaluation Quality and Validity
- 10.1** Results-Based Reporting and Logical Frameworks
- 10.2** The Two Components of a Program Theory Framework: Program Impact Models and Implementation Models
- 11.1** A More Detailed Look at the RealWorld Evaluation Design Frameworks
- 11.2** The RWE Approach to the Classification of Factors Affecting the Choice of Evaluation Design
- 11.3** The Strengths and Weaknesses of the Seven RWE Design Frameworks
- 11.4** Challenges Facing the Use of Experimental and Other Statistical Designs in RealWorld Evaluation Contexts
- 11.5** Examples of Randomized Control Trials
- 12.1** The Main Types of Questions That Can Be Included in Quantitative Surveys
- 12.2** Useful Sources of Secondary Data for QUANT Evaluations

- 12.3 Large-Scale Compilations of the Findings of Randomized Control Trials (RCTs)
- 12.4 Data Analysis for Quantitative Evaluations
- 14.1 Two Case Studies Illustrating Different Ways in Which Mixed-Method Designs Can Strengthen Impact Evaluations
- 14.2 Characteristics of Quantitative and Qualitative Approaches to Different Stages of the Evaluation Process
- 14.3 Common Issues Affecting the Validity of Statistical Impact Evaluation Designs and How Mixed-Method (MM) Designs Can Help Address Them
- 14.4 Three Case Studies Illustrating the Use of Mixed-Method Evaluations
- 15.1 Using Power Analysis and Effect Size for Estimating the Appropriate Sample Size for an Impact Evaluation

Specific Gender Evaluation Tools:

- 17.1 The Harvard Gender Analysis Framework
- 17.2 Tools for More In-Depth Gender-Responsive Evaluation (GRE) Designs
- 17.3 Two Examples of a Women's Empowerment Index

General Approaches and Methodologies for Gender Analysis:

- 17.4 The World Bank/IEG Implementation Completion Report: Gender Flag Review
- 17.5 Recommended Structure for the Evaluation Cooperation Group GRE Reports
- 17.6 Widely Used Gender Indices and Checklists
- 17.7 Evaluation Approaches Used in Standard GRE Designs
- 17.8 Examples of the Application of the Different GRE Designs
- 17.9 Evaluating the Gender Dimensions of Complex Development Programs
- 17.10 The Contribution of Gender-Responsive Budgeting (GRB) to National Policy Dialogue
- 17.11 Feminist Critical Theory
- 17.12 Gender-Sensitive Data-Collection Methods and Applications Used in the Case Studies in Appendix 17.14 and the Additional Time and Cost Implications (Compared to Standard Evaluation Methods)
- 17.13 Summary of Design and Data-Collection Methods Used in the Seven Case Studies Described in Appendix 17.14
- 17.14 Case Studies Illustrating Different Gender Impact Evaluation Methodologies
- 19.1 The Evaluation Framework and Scope of Work
- 19.2 The Actors Involved in an Evaluation of How Gender Equality Issues Were Addressed in an International Development Program
- 19.3 Procedures for Contracting Evaluation Consultants
- 19.4 Guidelines for Strengthening Terms of Reference
- 19.5 Building in Quality Assurance Procedures
- 19.6 Evaluation Capacity Development

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!

¹See 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 non-transparent 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 <i>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.</i>		
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 <i>RealWorld Evaluation</i> <i>For those who would like to dig deeper</i>		
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 indicators, 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 VanWyk 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.

¹Unless 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.

²Bold 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

³One 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.”

⁴For 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, socio-cultural, 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 technology⁵ 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)

⁵New 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)

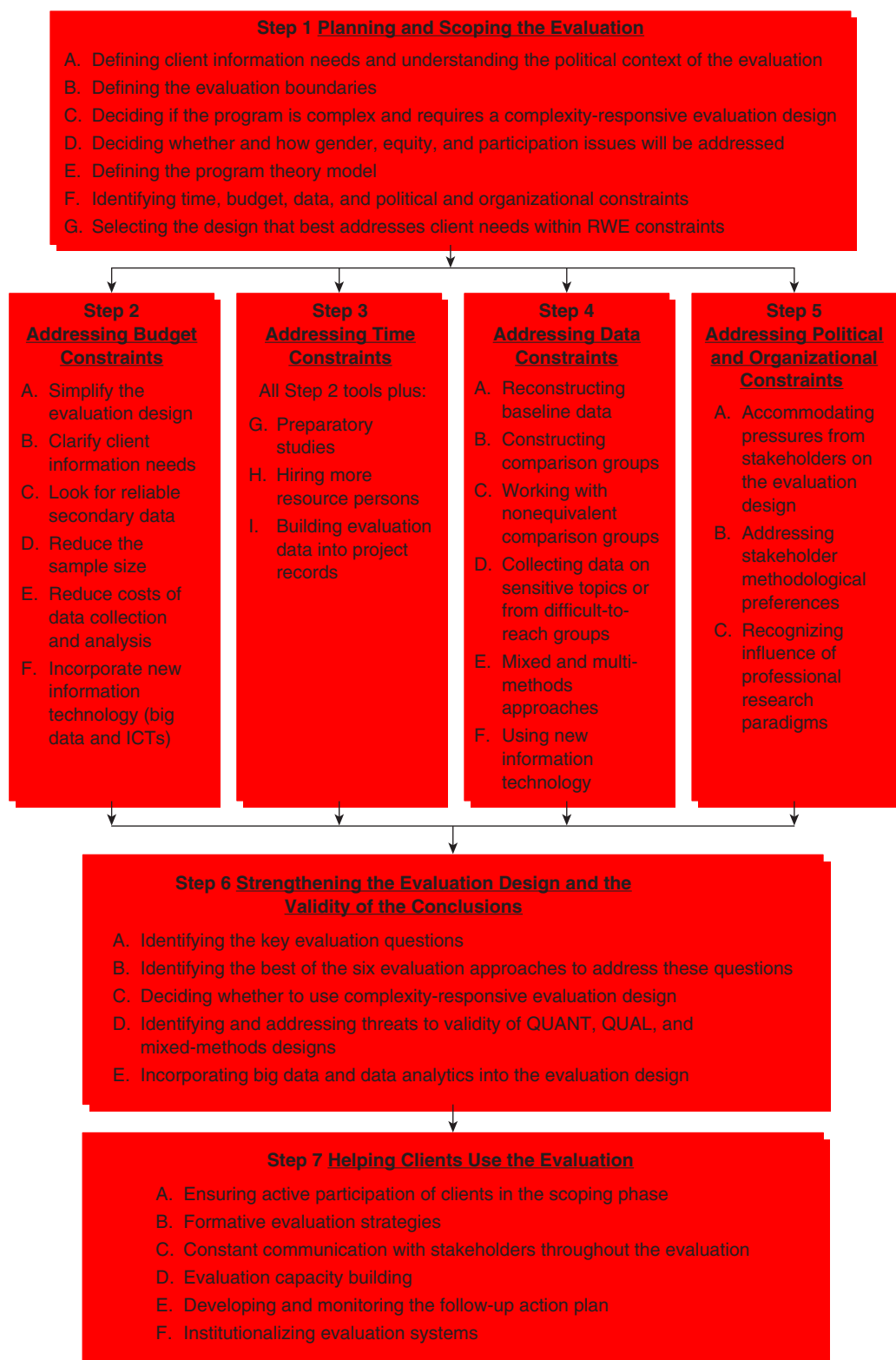
mobile phones and other handheld devices, and IOTs include devices for recording health, travel, and sensors attached to devices such as refrigerators or domestic and public electrical supply systems. However, these huge data sources are of limited practical utility without the use of smart data analytics 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-world 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.