

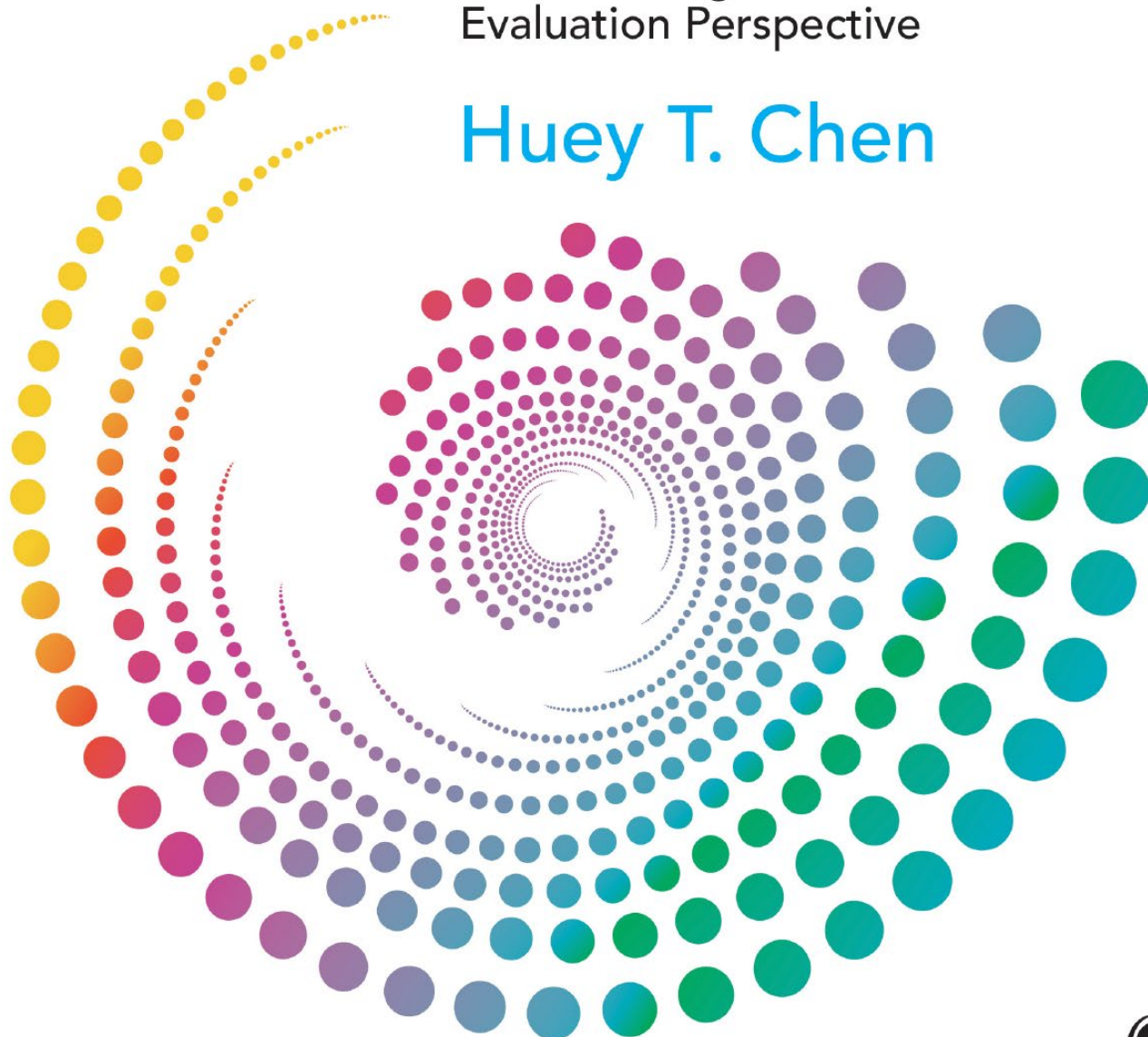
Second
Edition

A horizontal dotted line in a purple color, extending from the left edge of the cover towards the right, ending just before the main title.

PRACTICAL PROGRAM EVALUATION

Theory-Driven Evaluation
and the Integrated
Evaluation Perspective

Huey T. Chen



Second
Edition

**PRACTICAL
PROGRAM
EVALUATION**

To the memory of my mother, Huang-ai Chen

Second
Edition

PRACTICAL PROGRAM EVALUATION

Theory-Driven Evaluation
and the Integrated
Evaluation Perspective

Huey T. Chen

Mercer University



Los Angeles | London | New Delhi
Singapore | Washington DC



Los Angeles | London | New Delhi
Singapore | Washington DC

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 1/1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 048763

Acquisitions Editor: Helen Salmon
Associate Editor: Eve Oettinger
Editorial Assistant: Anna Villarruel
Production Editor: Jane Haenel
Copy Editor: Paula L. Fleming
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Susan Schon
Indexer: Robie Grant
Cover Designer: Anupama Krishnan
Marketing Manager: Nicole Elliott

Copyright © 2015 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Chen, Huey-tsyh.

Practical program evaluation : theory-driven evaluation and the integrated evaluation perspective / Huey T. Chen, Mercer University. — 2nd edition.

pages cm
Includes bibliographical references and index.

ISBN 978-1-4129-9230-5 (pbk. : alk. paper)

1. Evaluation research (Social action programs) I. Title.

H62.C3647 2015
300.72—dc23 2014019026

This book is printed on acid-free paper.

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

Contents

Preface	xvi
Special Features of the Book	xvii
About the Author	xix
PART I: INTRODUCTION	1
Chapter 1. Fundamentals of Program Evaluation	3
The Nature of Intervention Programs and Evaluation: A Systems View	3
Classic Evaluation Concepts, Theories, and Methodologies: Contributions and Beyond	6
Evaluation Typologies	7
<i>The Distinction Between Formative and Summative Evaluation</i>	7
<i>Analysis of the Formative and Summative Distinction</i>	8
<i>A Fundamental Evaluation Typology</i>	10
<i>Basic Evaluation Types</i>	11
<i>Hybrid Evaluation Types</i>	13
<i>Applications of the Fundamental Evaluation Typology</i>	14
Internal Versus External Evaluators	14
Politics, Social Justice, Evaluation Standards, and Ethics	15
Evaluation Steps	17
Evaluation Design and Its Components	18
Major Challenges of Evaluation:	
Lessons Learned From Past Practice	20
<i>Judge a Program Not Only by Its Results but Also by Its Context</i>	21
<i>Evaluations Must Address Both Scientific and Stakeholder Credibility</i>	22
<i>Evaluations Must Provide Information That Helps Stakeholders Do Better</i>	23
Addressing the Challenges: Theory-Driven Evaluation and the Integrated Evaluation Perspective	25

<i>Theory-Driven Evaluation Approach</i>	25
<i>Integrated Evaluation Perspective</i>	26
Program Complexity and Evaluation Theories	30
Who Should Read This Book and How They	
Should Use It	31
<i>Students</i>	31
<i>Evaluation Practitioners</i>	32
Introducing the Rest of the Chapters	32
Questions for Reflection	33

Chapter 2. Understand Approaches to

Evaluation and Select Ones That Work:

The Comprehensive Evaluation Typology	35
The Comprehensive Evaluation Typology: Means and Ends	36
<i>Stages in the Program Life Cycle</i>	39
<i>Dynamics of Transition Across Program Stages</i>	41
Evaluation Approaches Associated With Each Stage	41
<i>Planning Stage</i>	42
<i>Initial Implementation</i>	42
<i>Mature Implementation</i>	43
<i>Outcome Stage</i>	43
Strategies Underlying Evaluation Approaches	44
<i>Merit Assessment Strategies</i>	46
<i>Development Strategies</i>	47
Applying the Typology: Steps to Take	51
Evaluation Ranging Across Several Program Stages	54
Dynamics of Evaluation Entries Into Program Stages	55
1. <i>Single-Entry Evaluation</i>	55
2. <i>Multiple-Entry Evaluation</i>	56
Questions for Reflection	57

Chapter 3. Logic Models and the Action Model/Change

Model Schema (Program Theory)	58
Logic Models	58
<i>Additional Examples of Applying Logic Models</i>	61
Program Theory	65
The Action Model/Change Model Schema	66
<i>Descriptive Assumptions</i>	67
<i>Prescriptive Assumptions</i>	68
Components of the Change Model	70

<i>Goals and Outcomes</i>	71
<i>Determinants</i>	71
<i>Intervention or Treatment</i>	73
Components of the Action Model	74
<i>Intervention and Service Delivery Protocols</i>	74
<i>Implementing Organizations: Assess, Enhance, and Ensure Their Capabilities</i>	75
<i>Program Implementers: Recruit, Train, and Maintain Both Competency and Commitment</i>	76
<i>Associate Organizations/Community Partners: Establish Collaborations</i>	76
<i>Ecological Context: Seek the Support of the Environment</i>	77
<i>Target Population: Identify, Recruit, Screen, Serve</i>	78
Relationships Among Components of the Action Model/Change Model Schema	79
Applying the Action Model/Change Model Schema:	
An Example	82
<i>Change Model</i>	84
<i>Action Model</i>	84
Some Advantages of Using the Action Model/Change Model Schema	84
<i>Facilitation of Holistic Assessment</i>	84
<i>Provision of Comprehensive Information Needed to Improve Programs</i>	85
<i>Delineation of a Strategy to Consider Stakeholders' Views and Interests</i>	85
<i>Flexible Application of Research Methods to Serve Evaluation Needs</i>	86
<i>Aid to Selecting the Most Suitable Approaches or Methods</i>	86
Helping Stakeholders Gear Up (or Clear Up) Their Action Model/Change Model Schema	86
<i>Reviewing Existing Documents and Materials</i>	87
<i>Clarifying Stakeholders' Theory</i>	87
<i>Participatory Modes for Development Facilitation</i>	88
<i>Theorizing Procedures for Development Facilitation</i>	89
<i>Preparing a Rough Draft That Facilitates Discussion</i>	91
Applications of Logic Models and the Action Model/Change Model Schema	92
Questions for Reflection	92

PART II: PROGRAM EVALUATION TO HELP STAKEHOLDERS DEVELOP A PROGRAM PLAN	95
Chapter 4. Helping Stakeholders Clarify a Program Plan: Program Scope	97
The Program Plan, Program Scope, and Action Plan	97
Conceptual Framework of the Program Scope	98
<i>Why Develop a Program Scope?</i>	100
Strategies for Articulating the Program Scope	100
Background Information Provision Strategy and Approaches	101
<i>Needs Assessment</i>	101
<i>Formative Research</i>	102
The Conceptualization Facilitation Approach:	
A Part of the Development Facilitation Strategy	103
<i>Working Group or Intensive Interview Format?</i>	103
<i>Theorizing Methods</i>	104
<i>Determinants and Types of Action</i>	
<i>Model/Change Model Schemas</i>	109
<i>Choosing Interventions/Treatments That Affect the Determinant</i>	110
The Relevancy Testing Approach: A Part of the	
Troubleshooting Strategy	112
<i>Research Example of Relevancy Testing</i>	113
<i>Moving From Program Scope to Action Plan</i>	115
Questions for Reflection	116
Chapter 5. Helping Stakeholders Clarify a Program Plan: Action Plan	117
The Action Model Framework and the Action Plan	117
Strategies for Developing Action Plans	120
The Formative Research Approach	
(Under Background Information Provision Strategy)	120
<i>Example of Formative Research</i>	121
The Conceptualization Facilitation Approach	
(Under Development Facilitation Strategy)	122
1. <i>Implementing Organization: Assess, Enhance, and Ensure Its Capacity</i>	123
2. <i>Intervention and Service Delivery Protocols: Delineate Service Content and Delivery Procedures</i>	124
3. <i>Program Implementers: Recruit, Train, and Maintain for Competency and Commitment</i>	126
4. <i>Associate Organizations/Community Partners: Establish Collaborative Relationships</i>	127

5. Ecological Context: Seek the Support of the Environment	128
6. Target Population: Identify, Recruit, Screen, and Serve	130
Application of the Conceptualization Facilitation Strategy	133
<i>Example 1: A Garbage Reduction Program</i>	133
<i>Example 2: An HIV-Prevention Program</i>	136
The Pilot-Testing Approach	141
<i>Defining Pilot Testing</i>	141
<i>Conducting Pilot Testing</i>	142
<i>Designing Pilot Testing</i>	143
The Commentary or Advisory Approach	146
<i>Questions to Inform the Evaluator's Commentary</i> <i>on a Program Scope</i>	146
<i>Questions to Inform the Evaluator's Commentary</i> <i>on an Action Plan</i>	147
Summary	148
Questions for Reflection	149

PART III: EVALUATING IMPLEMENTATION 151

Chapter 6. Constructive Process Evaluation Tailored for the Initial Implementation	153
The Formative Evaluation Approach (Under the Troubleshooting Strategy)	154
<i>Timeliness and Relevancy</i>	155
<i>Research Methods</i>	155
<i>Steps in Applying Formative Evaluation</i>	156
<i>Four Types of Formative Evaluation</i>	158
<i>Formative Evaluation Results: Use With Caution</i>	164
The Program Review/Development Meeting (Under the Troubleshooting Strategy)	165
<i>Program Review/Development Meeting Principles</i> <i>and Procedures</i>	166
<i>Program Review/Development Meeting Advantages and</i> <i>Disadvantages</i>	167
<i>Example of a Program Review/Development Meeting</i>	169
Bilateral Empowerment Evaluation (Under the Development Partnership Strategy)	171
<i>Evaluation Process</i>	172
<i>Example of Bilateral Empowerment Evaluation</i>	172
<i>The Evaluator's Role</i>	173
<i>Pros and Cons of Bilateral Empowerment Evaluation</i>	174
Questions for Reflection	174

Chapter 7. Assessing Implementation in the	
Mature Implementation Stage	176
Constructive Process Evaluation and Its Application	177
<i>Modifying or Clarifying a Program Scope and Action Plan</i>	177
<i>Troubleshooting Implementation Problems</i>	179
Conclusive Process Evaluation and Its Applications	180
<i>How to Design a Conclusive Process Evaluation That Fits</i>	
<i>Stakeholders' Needs</i>	181
Approaches of Conclusive Process Evaluation	182
<i>Intervention Fidelity Evaluation</i>	183
<i>Referral Fidelity Evaluation</i>	184
<i>Service Delivery Fidelity Evaluation</i>	185
<i>Target Population Fidelity Evaluation</i>	186
<i>Fidelity Versus "Reinvention" in Conclusive</i>	
<i>Process Evaluation</i>	189
Hybrid Process Evaluation: Theory-Driven Process Evaluation	190
<i>Examples of Theory-Driven Process Evaluation</i>	191
<i>Theory-Driven Process Evaluation and</i>	
<i>Unintended Effects</i>	198
Questions for Reflection	199
 PART IV: PROGRAM MONITORING	
AND OUTCOME EVALUATION	201
 Chapter 8. Program Monitoring and the Development	
of a Monitoring System	203
What Is Program Monitoring?	203
Process Monitoring	204
<i>Uses of Process-Monitoring Data</i>	205
<i>Process Monitoring Versus Process Evaluation</i>	205
Outcome Monitoring	206
<i>Identification of Goals</i>	206
<i>Outcome Measures and Data Collection</i>	207
<i>Outcome Monitoring Versus Outcome Evaluation</i>	207
Program-Monitoring Systems Within Organizations	208
<i>Program-Monitoring System Elements</i>	209
<i>Developing a Program-Monitoring System</i>	210
 AN EXAMPLE OF DEVELOPING A PROGRAM	
MONITORING/EVALUATION SYSTEM	211
Questions for Reflection	229

Chapter 9. Constructive Outcome Evaluations	230
Constructive Outcome Evaluation	230
SMART Goals	231
<i>Specific</i>	232
<i>Measurable</i>	232
<i>Attainable</i>	232
<i>Relevant</i>	232
<i>Time-Bound</i>	233
<i>Putting SMART Characteristics Together in Goals</i>	233
Evaluability Assessment	234
<i>Step 1: Involve the Intended Users of Evaluation</i>	
<i>Information</i>	235
<i>Step 2: Clarify the Intended Program</i>	235
<i>Step 3: Explore the Program's Reality</i>	236
<i>Step 4: Reach Agreement on Any Needed Program</i>	
<i>Changes</i>	236
<i>Step 5: Explore Alternative Evaluation Designs</i>	236
<i>Step 6: Agree on the Evaluation's Priority and How</i>	
<i>Information From the Evaluation Will Be Used</i>	236
Plausibility Assessment/Consensus-Building Approach	237
<i>Potential Problems of Evaluation That Are Based Mainly</i>	
<i>on Official Goals</i>	237
<i>Plausibility Assessment/Consensus-Building Approach</i>	239
A Preview of Conclusive Outcome Evaluation: Selecting an	
Appropriate Approach	245
Questions for Reflection	246
 Chapter 10. The Experimentation Evaluation Approach	
to Outcome Evaluation	247
The Foundation of the Experimentation Approach	
to Outcome Evaluation	247
The Distinction Between Internal Validity and External	
Validity in the Campbellian Validity Typology	248
Threats to Internal Validity	249
Research Designs for Ruling Out Threats to Internal Validity	250
<i>Experimental Designs</i>	251
<i>Pre-Experimental Designs</i>	253
<i>Quasi-Experimental Designs</i>	256
Questions for Reflection	258

Chapter 11. The Holistic Effectuality Evaluation Approach	
to Outcome Evaluation	260
Ongoing Debates Over the Experimentation Evaluation Approach	260
Efficacy Evaluation Versus Effectiveness Evaluation	263
Relationships Among the Experimentation Evaluation	
Approach and the Campbellian Validity Typology	264
The Holistic Effectuality Evaluation Approach	266
<i>The Experimentation Evaluation Approach's</i>	
<i>Conceptualization of Outcome Evaluation</i>	267
<i>The Holistic Effectuality Approach's</i>	
<i>Conceptualization of Outcome Evaluation</i>	267
Constructive Assessment and Conclusive Assessment:	
Theory and Methodology	269
<i>Constructive Assessment</i>	270
<i>Conclusive Assessment</i>	275
THE ZUMBA WEIGHT LOSS PROJECT	275
Why Adjuvants Are Needed for Real-World Programs	281
<i>Types of Adjuvants and Threats to Internal Validity</i>	282
Methodology for Real-World Outcome Evaluation	284
<i>Assessing the Joint Effects of an Intervention</i>	
<i>and Its Adjuvants</i>	284
<i>Eliminating Potential Biases</i>	285
Research Steps for Assessing Real-World Effects	286
<i>Inquiring Into the Process of Contextualizing</i>	
<i>an Intervention in the Real-World Setting</i>	286
<i>Using a Relatively Unobtrusive Quantitative</i>	
<i>Design to Address Biases and Assess Change</i>	288
<i>Using an Auxiliary Design to Triangulate Evidence</i>	289
<i>Replication of the Mesological Intervention (Optional)</i>	291
EXAMPLE OF A REAL-WORLD OUTCOME EVALUATION	291
Checklist for Ranking Real-World Evaluations	296
Usefulness of the Holistic Effectuality Evaluation Approach	298
<i>Providing Theory and Methodology</i>	
<i>for Real-World Outcome Evaluation</i>	298
<i>Providing Insight Into the Relationship Between</i>	
<i>Adjuvants and Internal Validity</i>	298
<i>Inspiring Evaluators to Develop Indigenous Evaluation</i>	
<i>Theories and Methodologies</i>	299
The Experimentation Evaluation Approach Versus the	
Holistic Effectuality Evaluation Approach	299

<i>Pure Independent Effects</i>	300
<i>Real-World Joint Effects</i>	300
Building Evidence From the Ground Up	301
Questions for Reflection	302
Chapter 12. The Theory-Driven Approach to Outcome Evaluation	304
Clarifying Stakeholders' Implicit Theory	306
<i>Making Stakeholder Theory Explicit</i>	306
<i>Building Consensus Among Stakeholders</i>	
<i>Regarding Program Theory</i>	309
Guidelines for Conducting Theory-Driven Outcome Evaluation	309
Types of Theory-Driven Outcome Evaluation	310
The Intervening Mechanism Evaluation Approach	312
<i>Two Models of the Intervening Mechanism</i>	
<i>Evaluation Approach</i>	313
<i>Some Theoretical Bases of the Intervening</i>	
<i>Mechanism Evaluation</i>	318
<i>When to Use an Intervening Mechanism</i>	
<i>Evaluation Approach</i>	319
The Moderating Mechanism Evaluation Approach	321
<i>Constructing Moderating Mechanism Evaluation Models</i>	322
<i>Examples of Moderating Mechanism Evaluation</i>	323
<i>Advanced Moderating Mechanism Models</i>	324
<i>When to Use a Moderating Mechanism Evaluation Approach</i>	327
The Integrative Process/Outcome Evaluation Approach	327
<i>Research Methods and Strategies Associated With</i>	
<i>Integrative Process/Outcome Evaluation</i>	329
<i>Examples of Integrative Process/Outcome Evaluation</i>	329
Theory-Driven Outcome Evaluation and Unintended Effects	333
<i>Formal Specification of Possible Unintended Effects</i>	334
<i>Field Study Detection of Unintended Effects</i>	
<i>of Implementation</i>	335
A Reply to Criticisms of Theory-Driven Outcome Evaluation	336
Questions for Reflection	338
PART V: ADVANCED ISSUES IN PROGRAM EVALUATION	341
Chapter 13. What to Do if Your Logic Model Does	
Not Work as Well as Expected	343
A Diversity Enhancement Project	344
<i>Description of the Project</i>	344

<i>Applying the Logic Model</i>	345
<i>Applying the Action Model/Change Model Schema</i>	345
A Community Health Initiative	350
<i>Description of the Project</i>	350
<i>Applying the Logic Model</i>	352
<i>Applying the Action Model/Change Model Schema</i>	355
A Guide to Productively Applying the Logic Model and the Action Model/Change Model Schema	361
System Change and Evaluation in the Future	362
Questions for Reflection	363

Chapter 14. Formal Theories Versus Stakeholder Theories in Interventions: Relative Strengths and Limitations	365
Formal Theory Versus Stakeholder-Implicit Theory as a Basis for Intervention Programs	365
<i>Intervention Programs Based on Formal Theory</i>	365
<i>Programs Based on Stakeholder Theory</i>	366
<i>Views on the Relative Value of Formal Theory-Based Interventions and Stakeholder Theory-Based Interventions</i>	369
Formal Theory Versus Stakeholder Theory: A Case Study	371
<i>Program Theory Underlying the Anti-Secondhand Smoking Program</i>	371
<i>Action Model</i>	373
<i>Outcome Evaluation Design and Change Model</i>	376
<i>Process Evaluation Design</i>	377
Evaluation Findings	377
<i>Results of Process Evaluation</i>	377
<i>Outcome Evaluation Results</i>	378
Relative Strengths and Limitations of Formal Theory-Based Intervention and Stakeholder Theory-Based Intervention	379
<i>Theoretical Sophistication and Prior Evidence</i>	379
<i>Efforts to Clarify the Change Model and Action Model in Program Theory</i>	380
<i>Efficacious Evidence Versus Real-World Effectiveness</i>	380
<i>Viability</i>	381
<i>Action Theory Success and Conceptual Theory Success</i>	382
Lessons Learned From the Case Study	383
Questions for Reflection	387

Chapter 15. Evaluation and Dissemination: Top-Down	
Approach Versus Bottom-Up Approach	388
The Top-Down Approach to Transitioning From Evaluation to Dissemination	388
Lessons Learned From Applying the Top-Down Approach to Program Evaluation	389
Integrative Cogency Model: The Integrated Evaluation Perspective	394
<i>Effectual Cogency</i>	395
<i>Viable Cogency</i>	396
<i>Transferable Cogency</i>	398
Evaluation Approaches Related to the Integrative Cogency Model	398
<i>Effectuality Evaluation</i>	398
<i>Viability Evaluation</i>	399
<i>Transferability Evaluation</i>	399
The Bottom-Up Approach to Transitioning From Evaluation to Dissemination	400
<i>The Bottom-Up Approach</i>	400
<i>The Bottom-Up Approach and Social Betterment/Health Promotion Programs</i>	401
<i>Types of Intervention for the Bottom-Up Approach</i>	403
The Current Version of Evidence-Based Interventions: Limitations and Strategies to Address Them	403
The Integrated Evaluation Perspective on Concurrent Cogency Approaches	406
<i>Focusing on Effectual Cogency</i>	407
<i>Focusing on Viable Cogency</i>	407
<i>Optimizing Approach</i>	408
The Usefulness of the Bottom-Up Approach and the Integrative Cogency Model	408
Questions for Reflection	410
 References	 412
Index	426

Preface

I have been practicing program evaluation over a few decades. My practice has greatly benefited from conventional evaluation theories and approaches. However, on many occasions, I have also experienced conventional evaluation theories and approaches that do not work as well as they are supposed to. I have been contemplating and working on how to expand them or develop alternative theories and approaches that will better serve evaluation in the future. I planned to discuss my experiences and lessons learned from these efforts in the second edition of *Practical Program Evaluation* so that evaluators, new or seasoned, would not only learn both traditional and cutting-edge concepts but also have opportunities to participate in further advancing program evaluation. However, this plan has frequently been stymied. One reason is that the more I study the issues, the more complicated they become. I sometimes felt as though I was constantly banging my head against the proverbial wall. Luckily, I found I was not the only person having these frustrations and struggling with these problems. The following friends and colleagues have provided timely encouragement and advice that have been crucial to my finishing the book: Thomas Chapel, Amy DeGroff, Stewart Donaldson, Jennifer Greene, Brian Lien, Lorine Spencer, Jonathan Morell, Craig Thomas, Nannette Turner, and Jennifer Urban. I am indebted greatly to them for their support of the project. I am also grateful for the valuable feedback from the following reviewers: Darnell J. Bradley, Cardinal Stritch University; C. W. Cowles, Central Michigan University; and Mario A. Rivera, University of New Mexico. Any shortcomings of this book are entirely my own.

Furthermore, the book was also frequently disrupted by other, more pressing tasks. Helen Salmon, my SAGE editor, issued gentle ongoing reminders and patiently checked on my progress every step of the way. Without her persistent nudging, I would not have been able to meet the deadline. I also appreciate my research assistants, Joanna Hill and Mauricia Barnett, for their help in preparing questions for reflection and the tables that appear in the book. With so much time and effort spent, it is a great joy for me to see this book reach fruition.

Special Features of the Book

This book is about program evaluation in action, and to that end it does the following:

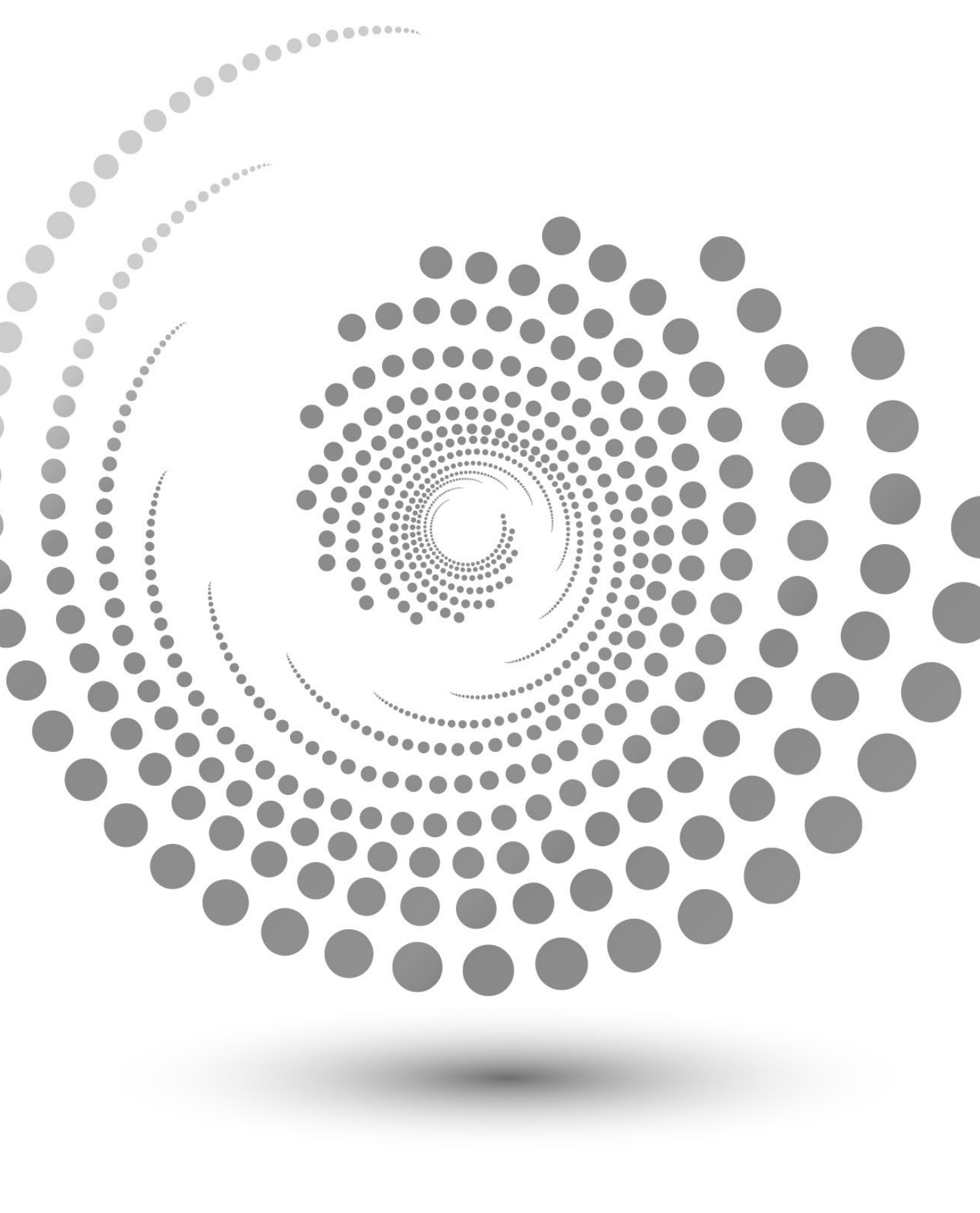
1. *Provides a comprehensive evaluation typology that facilitates the systematic identification of stakeholders' needs and the selection of the evaluation options best suited to meet those needs.* Almost always, program evaluation is initiated to meet the particular evaluation needs of a program's stakeholders. If a program evaluation is to be useful to those stakeholders, it is their expectations that evaluators must keep in mind when designing the evaluation. The precise communication and comprehension of stakeholder expectations is crucial; to facilitate the communication process, this book presents a comprehensive evaluation typology for the effective identification of evaluation needs. Within this typology, the book provides a variety of evaluation approaches suitable across a program's life cycle—from program planning to initial implementation, mature implementation, and outcome achievement—to enrich the evaluator's toolbox. Once the stakeholders' expectations are identified, evaluators must select a strategy for addressing each evaluation need. Many evaluation options are available. The book discusses them, exploring the pros and cons of each and acknowledging that trade-offs sometimes must be made. Furthermore, it suggests practical principles that can guide evaluators to make the best choices in the evaluation situations they are likely to encounter.

2. *Introduces both conventional and cutting-edge evaluation perspectives and approaches.* The core of program evaluation is its body of concepts, theories, and methods. It provides evaluators needed principles, strategies, and tools for conducting evaluations. As will be demonstrated in the book, cutting-edge evaluation approaches have been developed to further advance program evaluation by thinking outside the proverbial box. Evaluators can do better evaluations if they are familiar and competent with both conventional and innovative evaluation perspectives and approaches. This book systematically introduces the range of options and discusses the conditions under which they can be fruitfully applied.

3. *Puts each approach into action.* Using illustrative examples from the field, the book details the methods and procedures involved in using various evaluation options. How does the program evaluator carry out an evaluation so as to meet real evaluation needs? Here, practical approaches are discussed—yet this book avoids becoming a “cookbook.” The principles and strategies of evaluation that it presents are backed by theoretical justifications, which are also explained. This context, it is hoped, fosters the latitude, knowledge, and flexibility with which program evaluators can design suitable evaluation models for a particular evaluation project and better serve stakeholders’ needs.

About the Author

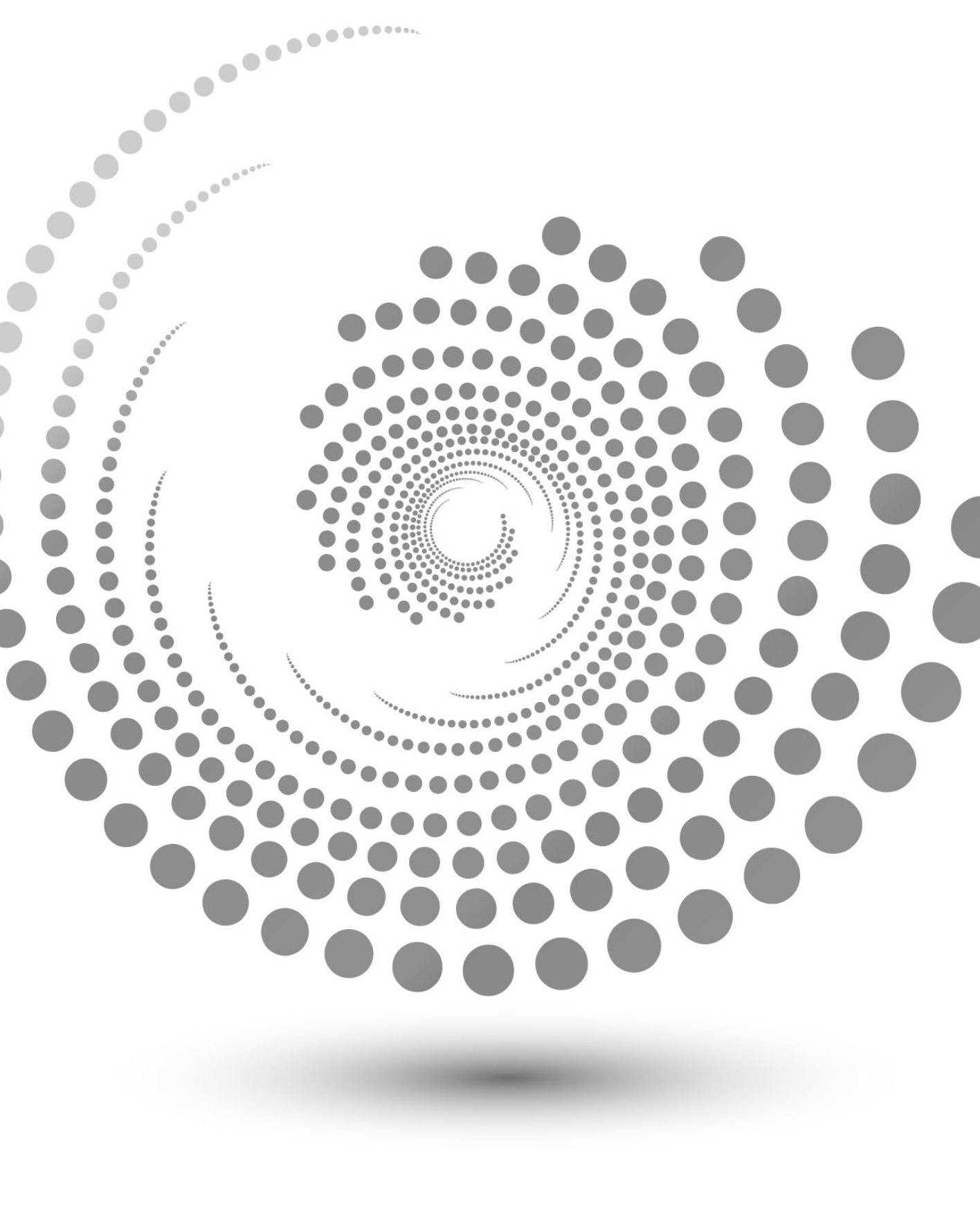
Huey T. Chen is Professor of the Department of Public Health and Director of the Center for Evaluation and Applied Research in the College of Health Professions at Mercer University. He previously served as branch chief and senior evaluation scientist at the Centers for Disease Control and Prevention (CDC), as well as Professor at the University of Alabama at Birmingham. Dr. Chen has worked with community organizations, health-related agencies, government agencies, and educational institutions. He has conducted both large-scale and small-scale evaluations in the United States and internationally, including evaluating a drug abuse treatment program and a youth service program in Ohio, a carbon monoxide ordinance in North Carolina, a community health initiative in New Jersey, a juvenile delinquency prevention and treatment policy in Taiwan, and an HIV prevention and care initiative in China. He has written extensively on program theory, theory-driven evaluation, the bottom-up evaluation approach, and the integrated evaluation perspective. In addition to publishing over 70 articles in peer-reviewed journals, he is the author of several evaluation books. His book *Theory-Driven Evaluations* (1990, SAGE) is seen as one of the landmarks in program evaluation. His book *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective, Second Edition* (2015, SAGE) introduces cutting-edge evaluation approaches and illustrates the benefits of thinking outside the proverbial box. Dr. Chen serves on the editorial advisory boards of *Evaluation and Program Planning* and is a winner of the American Evaluation Association's Lazarsfeld Award for Evaluation Theory and of the Senior Biomedical Service Award from the CDC for his evaluation work.



PART I

Introduction

The first three chapters of this book, which comprise Part I, provide general information about the theoretical foundations and applications of program evaluation principles. Basic ideas are introduced, and a conceptual framework is presented. The first chapter explains the purpose of the book and discusses the nature, characteristics, and strategies of program evaluation. In Chapter 2, program evaluators will find a systematic typology of the various evaluation approaches one can choose among when faced with particular evaluation needs. Chapter 3 introduces the concepts of logic models and program theory, which underlie many of the guidelines found throughout the book.



CHAPTER 1

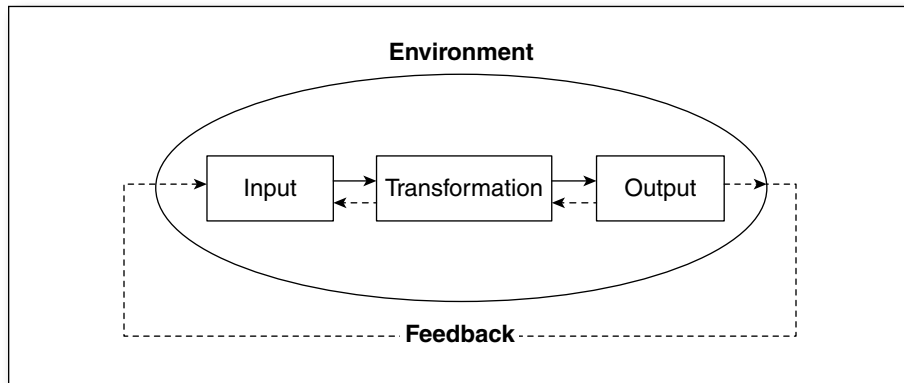


Fundamentals of Program Evaluation

The programs that evaluators can expect to assess have different names such as *treatment* program, *action* program, or *intervention* program. These programs come from different substantive areas, such as health promotion and care, education, criminal justice, welfare, job training, community development, and poverty relief. Nevertheless, they all have in common organized efforts to enhance human well-being—whether by preventing disease, reducing poverty, reducing crime, or teaching knowledge and skills. For convenience, programs and policies of any type are usually referred in this book as “intervention programs” or simply “programs.” An intervention program intends to change individuals’ or groups’ knowledge, attitudes, or behaviors in a community or society. Sometimes, an intervention program aims at changing the entire population of a community; this kind of program is called a *population-based intervention program*.

THE NATURE OF INTERVENTION PROGRAMS AND EVALUATION: A SYSTEMS VIEW

The terminology of systems theory (see, e.g., Bertalanffy, 1968; Ryan & Bohman, 1998) provides a useful means of illustrating how an intervention program works as an open system, as well as how program evaluation serves the program. In a general sense, as an open system an intervention program consists of five components (input, transformation, outputs, environment, and feedback), as illustrated in Figure 1.1.

Figure 1.1 A Systems View of a Program

Inputs. Inputs are resources the program takes in from the environment. They may include funding, technology, equipment, facilities, personnel, and clients. Inputs form and sustain a program, but they cannot work effectively without systematic organization. Usually, a program requires an implementing organization that can secure and manage its inputs.

Transformation. A program converts inputs into outputs through transformation. This process, which begins with the initial implementation of the treatment/intervention prescribed by a program, can be described as the stage during which implementers provide services to clients. For example, the implementation of a new curriculum in a school may mean the process of teachers teaching students new subject material in accordance with existing instructional rules and administrative guidelines. Transformation also includes those sequential events necessary to achieve desirable outputs. For example, to increase students' math and reading scores, an education program may need to first boost students' motivation to learn.

Outputs. These are the results of transformation. One crucial output is the attainment of the program's goals, which justifies the existence of the program. For example, an output of a treatment program directed at individuals who engage in spousal abuse is the end of the abuse.

Environment. The environment consists of any factors that, despite lying outside a program's boundaries, can nevertheless either foster or constrain that program's implementation. Such factors may include social norms, political structures, the economy, funding agencies, interest groups, and concerned

citizens. Because an intervention program is an open system, it depends on the environment for its inputs: clients, personnel, money, and so on. Furthermore, the continuation of a program often depends on how the general environment reacts to program outputs. Are the outputs valuable? Are they acceptable? For example, if the staff of a day care program is suspected of abusing children, the environment would find that output unacceptable. Parents would immediately remove their children from the program, law enforcement might press criminal charges, and the community might boycott the day care center. Finally, the effectiveness of an open system, such as an intervention program, is influenced by external factors such as cultural norms and economic, social, and political conditions. A contrasting system may be illustrative: In a biological system, the use of a medicine to cure an illness is unlikely to be directly influenced by external factors such as race, culture, social norms, or poverty.

Feedback. So that decision makers can maintain success and correct any problems, an open system requires information about inputs and outputs, transformation, and the environment's responses to these components. This feedback is the basis of program evaluation. Decision makers need information to gauge whether inputs are adequate and organized, interventions are implemented appropriately, target groups are being reached, and clients are receiving quality services. Feedback is also critical to evaluating whether outputs are in alignment with the program's goals and are meeting the expectations of stakeholders. Stakeholders are people who have a vested interest in a program and are likely be affected by evaluation results; they include funding agencies, decision makers, clients, program managers, and staff. Without feedback, a system is bound to deteriorate and eventually die. Insightful program evaluation helps to both sustain a program and prevent it from failing. The action of feedback within the system is indicated by the dotted lines in Figure 1.1.

To survive and thrive within an open system, a program must perform at least two major functions. First, internally, it must ensure the smooth transformation of inputs into desirable outcomes. For example, an education program would experience negative side effects if faced with disruptions like high staff turnover, excessive student absenteeism, or insufficient textbooks. Second, externally, a program must continuously interact with its environment in order to obtain the resources and support necessary for its survival. That same education program would become quite vulnerable if support from parents and school administrators disappeared.

Thus, because programs are subject to the influence of their environment, every program is an open system. The characteristics of an open system can also be identified in any given policy, which is a concept closely related to that of a program. Although policies may seem grander than programs—in terms of

the envisioned magnitude of an intervention, the number of people affected, and the legislative process—the principles and issues this book addresses are relevant to both. Throughout the rest of the book, the word *program* may be understood to mean program *or* policy.

Based upon the above discussion, this book defines *program evaluation* as the process of systematically gathering empirical data and contextual information about an intervention program—specifically answers to what, who, how, whether, and why questions that will assist in assessing a program’s planning, implementation, and/or effectiveness. This definition suggests many potential questions for evaluators to ask during an evaluation: The “what” questions include those such as, what are the intervention, outcomes, and other major components? The “who” questions might be, who are the implementers and who are the target clients? The “how” questions might include, how is the program implemented? The “whether” questions might ask whether the program plan is sound, the implementation adequate, and the intervention effective. And the “why” questions could be, why does the program work or not work? One of the essential tasks for evaluators is to figure out which questions are important and interesting to stakeholders and which evaluation approaches are available for evaluators to use in answering the questions. These topics will be systematically discussed in Chapter 2. The purpose of program evaluation is to make the program accountable to its funding agencies, decision makers, or other stakeholders and to enable program management and implementers to improve the program’s delivery of acceptable outcomes.

CLASSIC EVALUATION CONCEPTS, THEORIES, AND METHODOLOGIES: CONTRIBUTIONS AND BEYOND

Program evaluation is a young applied science; it began developing as a discipline only in the 1960s. Its basic concepts, theories, and methodologies have been developed by a number of pioneers (Alkin, 2013; Shadish, Cook, & Leviton, 1991). Their ideas, which are foundational knowledge for evaluators, guide the design and conduct of evaluations. These concepts are commonly introduced to readers in two ways. The conventional way is to introduce classic concepts, theories, and methodologies exactly as proposed by these pioneers. Most major evaluation textbooks use this popular approach.

This book, however, not only introduces these classic concepts, theories, and methodologies but also demonstrates how to use them as a foundation for formulating additional evaluation approaches. Readers can not only learn from evaluation pioneers’ contributions but also expand or extend their work, informed by lessons learned from experience or new developments in program evaluation. However, there is a potential drawback to taking this path. It

requires discussing the strengths and limitations of the work of the field's pioneers. Such critiques may be regarded as intended to diminish or discredit this earlier work. It is important to note that the author has greatly benefited from the classic works in the field's literature and is very grateful for the contributions of those who developed program evaluation as a discipline. Moreover, the author believes that these pioneers would be delighted to see future evaluators follow in their footsteps and use their accomplishments as a basis for exploring new territory. In fact, the seminal authors in the field would be very upset if they saw future evaluators still working with the same ideas, without making progress. It is in this spirit that the author critiques the literature of the field, hoping to inspire future evaluators to further advance program evaluation.

Indeed, the extension or expansion of understanding is essential for advancing program evaluation. Readers will be stimulated to become independent thinkers and feel challenged to creatively apply evaluation knowledge in their work. Students and practitioners who read this book will gain insights from the discussions of different options, formulate their own views of the relative worth of these options, and perform better work as they go forward in their careers.

EVALUATION TYPOLOGIES

Stakeholders need two kinds of feedback from evaluation. The first kind is information they can use to improve a program. Evaluations can function as improvement-oriented assessments that help stakeholders understand whether a program is running smoothly, whether there are problems that need to be fixed, and how to make the program more efficient or more effective. The second kind of feedback evaluations can provide is an accountability-oriented assessment of whether or not a program has worked. This information is essential for program managers and staff to fulfill their obligation to be accountable to various stakeholders.

Different styles of evaluation have been developed to serve these two types of feedback. This section will first discuss Scriven's (1967) classic distinction between formative and summative evaluation and then introduce a broader evaluation typology.

The Distinction Between Formative and Summative Evaluation

Scriven (1967) made a crucial contribution to evaluation by introducing the distinction between formative and summative evaluation. According to Scriven, formative evaluation fosters improvement of ongoing activities. Summative evaluation, on the other hand, is used to assess whether results have met the stated goals.

Summative evaluation informs the go or no-go decision, that is, whether to continue or repeat a program or not. Scriven initially developed this distinction from his experience of curriculum assessment. He viewed the role of formative evaluation in relation to the ongoing improvement of the curriculum, while the role of summative evaluation serves administrators by assessing the entire finished curriculum. Scriven (1991a) provided more elaborated descriptions of the distinction. He defined *formative evaluation* as “evaluation designed, done, and intended to support the process of improvement, and normally commissioned or done, and delivered to someone who can make improvement” (p. 20). In the same article, he defined *summative evaluation* as “the rest of evaluation; in terms of intentions, it is evaluation done for, or by, any observers or decision makers (by contrast with developers) who need valutive conclusions for any other reasons besides development.” The distinct purposes of these two kinds of evaluation have played an important role in the way that evaluators communicate evaluation results to stakeholders.

Scriven (1991a) indicated that the best illustration of the distinction between formative and summative evaluation is the analogy given by Robert Stake: “When the cook tastes the soup, that’s formative evaluation; when the guest tastes it, that’s summative evaluation” (Scriven, p. 19). The cook tastes the soup while it is cooking in case, for example, it needs more salt. Hence, formative evaluation happens in the early stages of a program so the program can be improved as needed. On the other hand, the guest tastes the soup after it has finished cooking and is served. The cook could use the guest’s opinion to determine whether to serve the soup to other guests in the future. Hence, summative evaluation happens in the last stage of a program and emphasizes the program’s outcome.

Scriven (1967) placed a high priority on summative evaluation. He argued that decision makers can use summative evaluation to eliminate ineffective programs and avoid wasting money. However, Cronbach (1982) disagreed with Scriven’s view, arguing that program evaluation is most useful when it provides information that can be used to strengthen a program. He also implied that few evaluation results are used for making go or no-go decisions. Which type of evaluation has a higher priority is an important issue for evaluators, and the importance of this issue will be revisited later in this chapter.

Analysis of the Formative and Summative Distinction

The distinction between formative and summative evaluation provides an important framework evaluators can use to communicate ideas and develop approaches, and these concepts will continue to play an important role. However, Scriven (1991a) proposed that formative and summative evaluations are the two main evaluation types. In reality, there are other important evaluation types that are not

covered in this distinction. To avoid confusion and to lay a foundation for advancing the discipline, it is important to highlight these other evaluation types as well.

In Scriven's conceptualization, evaluation serves to improve a program only during earlier stages of the program (formative evaluation), while evaluation renders a final verdict at the outcome stage (summative evaluation). However, this conceptualization may not sufficiently cover many important evaluation activities (Chen, 1996). For example, evaluations at the early stage of the program do not need to be used to improve the program. Evaluators could administer summative evaluations during earlier phases of the program. Similarly, evaluations conducted at the outcome stage do not have to be summative. Evaluators could administer a formative evaluation at the outcome stage to gain information that would inform and improve future efforts.

Since Scriven regarded Robert Stake's soup-tasting analogy as the best way to illustrate the formative/summative distinction, let's use this analogy to illustrate that all evaluations do not fit this description. According to Stake's analogy, when "the cook tastes the soup," that act represents formative evaluation. This concept of formative evaluation has some limitations. The cook does not always taste the soup for the purpose of improvement. The cook may taste the soup to determine whether the soup is good enough to serve to the guests at all, especially if it is a new recipe. Upon testing the soup, she/he may feel it is good enough to serve to the guests; alternatively, she/he may decide that the soup is awful and not worth improving and simply chuck the soup and scratch it off the menu. In this case, the cook has not tasted the soup for the purpose of improvement but to reach a conclusion about including the soup or excluding it from the menu.

To give another illustration, a Chinese cook, who is a friend of mine, once tried to prepare a new and difficult dish, called Peking duck, for his restaurant. Tasting his product, he found that the skin of the duck was not as crispy as it was supposed to be, nor the meat as flavorful. Convinced that Peking duck was beyond his capability as a chef, he decided not to prepare the dish again. Again, the cook tasted the product to conduct a summative assessment rather than a formative one. The formative/summative distinction does not cover this kind of evaluation.

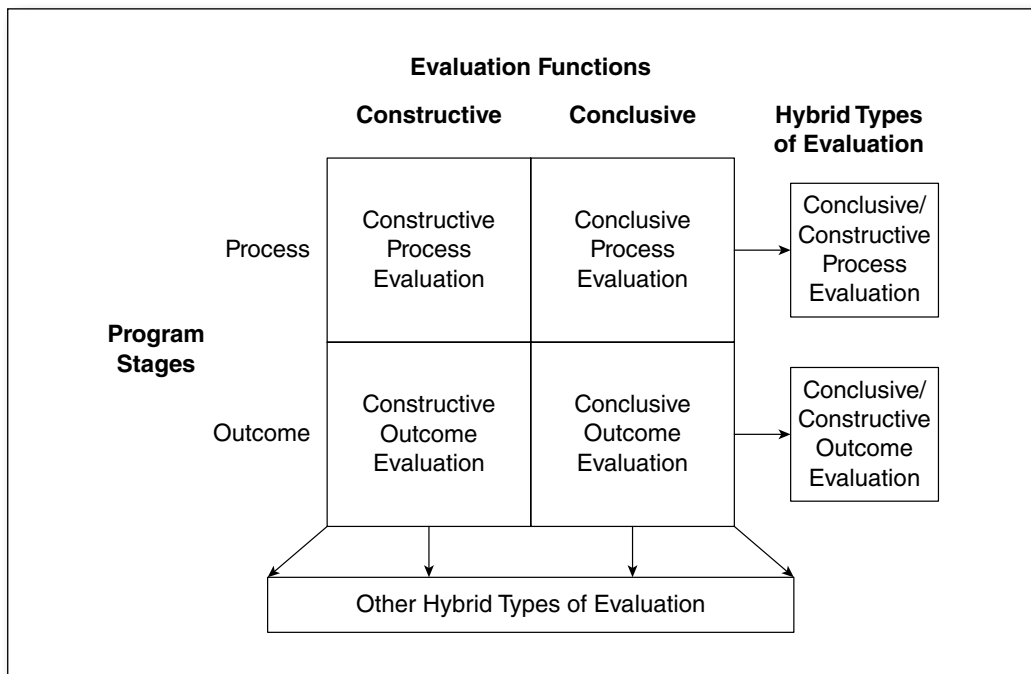
Returning to Stake's analogy, when "the guest tastes the soup," this is regarded as a summative evaluation since the guest provides a conclusive opinion of the soup. This concept of summative evaluation also has limitations. For example, the opinion of the guests is not always used solely to determine the soup's final merit. Indeed, a cook might well elicit opinions from the guests for the purpose of improving the soup in the future. In this case, this type of evaluation is also not covered by the formative/summative distinction.

Stake's analogy, though compelling, excludes many evaluation activities. Thus, we need a broader conceptual typology so as to more comprehensively communicate or guide evaluation activities.

A Fundamental Evaluation Typology

To include more evaluation types in the language used to communicate and guide evaluation activities, this chapter proposes to extend Scriven's formative and summative distinction. The typology developed here is a reformulation of an early work by Chen (1996). This typology has two dimensions: the program stages and evaluation functions. In terms of program stages, evaluation can focus on program process (such as program implementation) and/or on program outcome (such as the impact of the program on its clients). In terms of evaluation functions, evaluation can serve a constructive function (providing information for improving a program) and/or a conclusive function (judging the overall merit or worth of a program). A fundamental typology of evaluation can thus be developed by placing program stages and evaluation functions in a matrix, as shown in Figure 1.2.

Figure 1.2 Fundamental Evaluation Typology



SOURCE: Adapted from Chen (1996).

This typology consists of both basic evaluation types and hybrid evaluation types. The rest of this section will discuss the basic types first and then the hybrid types.

Basic Evaluation Types

The basic types of evaluation include constructive process evaluation, conclusive process evaluation, constructive outcome evaluation, and conclusive outcome evaluation.

Constructive Process Evaluation

Constructive process evaluation provides information about the relative strengths/weaknesses of the program's structure or implementation processes, with the purpose of program improvement. Constructive process evaluation usually does not provide an overall assessment of the success or failure of program implementation. For example, a constructive process evaluation of a family-planning program may indicate that more married couples can be persuaded to utilize birth control in an underdeveloped country if the service providers or counselors are local people, rather than outside health workers. This information does not provide a conclusive judgment of the merits of program implementation, but it is useful for improving the program. Decision makers and program designers can use the information to strengthen the program by training more local people to become service providers or counselors.

Conclusive Process Evaluation

This type of evaluation, which is frequently used, is conducted to judge the merits of the implementation process. Unlike constructive process evaluation, conclusive process evaluation attempts to judge whether the implementation of a program is a success or a failure, appropriate or inappropriate. A good example of conclusive process evaluation is an assessment of whether program services are being provided to the target population. If an educational program intended to serve disadvantaged children is found to serve middle-class children, the program would be considered an implementation failure. Another good example of conclusive process evaluation is manufacturing quality control, when a product is rejected if it fails to meet certain criteria. Vivid examples of conclusive process evaluation are the investigative reports seen on popular TV programs, such as *60 Minutes* and *20/20*. In these programs, reporters use hidden cameras to document

whether services delivered by such places as psychiatric hospitals, nursing homes, child care centers, restaurants, and auto repair shops are appropriate.

Constructive Outcome Evaluation

This type of evaluation identifies the relative strengths and/or weaknesses of program elements in terms of how they may affect program outcomes. This information can be useful for improving the degree to which a program is achieving its goals, but it does not provide an overall judgment of program effectiveness. For example, evaluators may facilitate a discussion among stakeholders to develop a set of measurable goals or to reach consensus about program goals. Again, such activity is useful for improving the program's chance of success, but it stops short of judging the overall effectiveness of the program. This type of evaluation will be discussed in detail in Chapter 9. In another example, a service agency may have two types of social workers, case managers whose work is highly labor-intensive and care managers whose work is less labor-intensive. An evaluator can apply constructive outcome evaluation to determine which kind of social worker is more cost-effective for the agency.

Conclusive Outcome Evaluation

The purpose of a conclusive outcome evaluation is to provide an overall judgment of a program in terms of its merit or worth. Scriven's summative evaluation is synonymous with this category. A typical example of conclusive outcome evaluation is validity-focused outcome evaluation that determines whether changes in outcomes can be causally attributed to the program's intervention. This kind of evaluation is discussed in detail in Chapter 10.

The typology outlined above eliminates some of the difficulties found in the soup-tasting analogy. Formerly, when the cook tasted the soup for conclusive judgment purposes, this activity did not fit into the formative/summative distinction. However, it can now be classified as conclusive process evaluation. Similarly, when the guest tastes the soup for improvement purposes, this action can now be classified as constructive outcome evaluation.

Furthermore, the typology clarifies the myth that process evaluation is always a kinder, gentler type of evaluation in which evaluators do not make tough conclusive judgments about the program. Constructive process evaluation may be kinder and gentler, but conclusive process evaluation is not necessarily so. For example, TV investigative reports that expose the wrongdoing in a psychiatric hospital, auto shop, restaurant, or day care center have resulted in changes in service delivery, the firing of managers and employees, and even the closing of

the agencies or businesses in question. In such cases, process evaluations were tougher than many outcome evaluations in terms of critical assessment and impact. Moreover, the basic typology disrupts the notion that outcome evaluation must always be carried out with a “macho” attitude so that it threatens program providers while failing to offer any information about the program. A conclusive outcome evaluation may provide information whether a program has been successful or not, but the constructive outcome evaluation can provide useful information for enhancing the effectiveness of a program without threatening its existence. For example, the survival of a program is not threatened by a constructive outcome evaluation that indicates that program effectiveness could be improved by modifying some intervention elements or procedures.

Hybrid Evaluation Types

Another important contribution of this fundamental evaluation typology is to point out that evaluators can move beyond the basic evaluation types to conduct hybrid evaluations. As illustrated in Figure 1.2, a hybrid evaluation can combine evaluation functions, program stages, or both (Chen, 1996). This section intends to introduce two types of hybrid evaluation that, across evaluation, functions at a program stage.

Conclusive/Constructive Process Evaluation

Conclusive/constructive process evaluation serves both accountability and program improvement functions. A good example is evaluation carried out by the Occupational Safety and Health Administration (OSHA). OSHA inspectors may evaluate a factory to determine whether the factory passes a checklist of safety and health rules and regulations. The checklist is so specific, however, that these inspections can also be used for improvement. If a company fails the inspection, the inspector provides information concerning areas that need correction to satisfy safety standards. Other regulatory agencies, such as the Environmental Protection Agency (EPA), perform a similar type of evaluation. In these kinds of evaluation, the overall quality of implementation is represented by a checklist of crucial elements. These elements provide exact clues for how to comply with governmental regulations.

A similar principle can be applied to assess the implementation of an intervention. As will be discussed in Chapter 7, a conclusive/constructive process evaluation can look into both overall quality and discrete program elements so as to provide information about the overall quality of implementation as well as specific areas for its future improvement.

Conclusive/Constructive Outcome Evaluation

Another hybrid evaluation type is the conclusive/constructive outcome evaluation. An excellent example of this kind of evaluation is real-world outcome evaluation, which will be discussed in great detail in Chapter 11. Another excellent example is theory-driven outcome evaluation. This type of evaluation elaborates causal mechanisms underlying a program so that it examines not only whether the program has an impact but why. It also informs stakeholders as to which mechanisms influence program success or failure for program improvement purposes. Theory-driven outcome evaluation will be discussed in Chapters 12 and 14 of the book.

Applications of the Fundamental Evaluation Typology

The fundamental evaluation typology discussed here prevents evaluators from hewing rigidly to just two types of evaluation, that is, formative evaluation in the early stages of the program and summative evaluation toward the end. The fundamental evaluation typology provides evaluators and stakeholders many options for devising basic or hybrid types of evaluation at implementation and outcome stages so as to best meet stakeholders' needs. However, the fundamental evaluation typology does not cover the planning stage. Thus, Chapter 2 will expand the fundamental evaluation typology into a comprehensive evaluation typology that covers a full program cycle from program planning to implementation to outcome. Then the rest of the book will provide concrete examples of these evaluation approaches and illustrate their applications across the entire life cycle of programs.

INTERNAL VERSUS EXTERNAL EVALUATORS

Evaluators are usually classified into two categories: internal and external evaluators. Internal evaluators are employed by an organization and are responsible for evaluating the organization's own programs. External evaluators are not employees of the organization but are experts hired from outside to evaluate the program. One of the major differences between the two is independence. Internal evaluators are part of the organization. They are familiar with the organizational culture and the programs to be evaluated. Like other employees, they share a stake in the success of the organization. External evaluators are not constrained by organizational management and relationships with staff members and are less invested in the program's success. The general conditions that tend to favor either internal evaluation or external evaluation are summarized as follows:

Internal Evaluation

- Cost is a great concern.
- Internal capacity/resources are available.
- The evaluator's familiarity with the program is important.
- The program is straightforward.
- Evaluation is for the purpose of monitoring or is constructive in nature.

External Evaluation

- The cost of hiring an external evaluator is manageable.
- Independence and objectivity are essential.
- A program is large or complicated.
- The evaluation will focus on conclusive assessment or conclusive/constructive assessment.
- Comprehensive assessment or fresh insight is needed.

**POLITICS, SOCIAL JUSTICE, EVALUATION
STANDARDS, AND ETHICS**

One important distinction that separates program evaluation from research is that evaluations are carried out under political processes. The purpose of an evaluation is to evaluate an intervention program. However, the program is created by political processes. What kinds of programs are to be funded? Which programs need evaluation in a community? These decisions are made through bargaining and negotiation by key players such as politicians and advocacy groups. After a program is funded and evaluators are hired to evaluate it, the focus of the evaluation and the questions to be asked are determined, or largely influenced, by stakeholders. Cronbach and colleagues (1980) argued that a theory of evaluation must be as much a theory of political interaction as it is a theory of how to determine facts. Weiss (1998), too, indicated that evaluators must understand the political nature of evaluations and be aware of the obstacles and opportunities that can impinge upon evaluation efforts.

Since evaluation provides feedback to a program, evaluators may have high hopes that decision makers will use the findings as a basis for action. However, since program evaluation is part of political processes, evaluation findings are just one of many inputs that decision makers use. Decision making is more often based on factors such as political support and community service needs than evaluation findings. Since evaluations take place within a political and an organizational context, Chelimsky (1987) stated that evaluators are shifting their view of the role evaluations play, from reforming society to the more realistic aim of bringing the

best possible information to bear on a wide variety of policy questions. Also because evaluation takes place in a political environment, evaluators' communication skills are critical. Evaluators' qualifications should include research skills but should emphasize group facilitation skills, political adroitness, managerial ability, and cultural sensitivity to multiple stakeholders.

In evaluation, stakeholders are those persons, groups, or organizations who have a vested interest in the evaluation results. Stakeholders often are not a homogeneous group but rather multiple groups with different interests, priorities, and degrees of power or influence. The number of stakeholder groups evaluators must communicate with often depends on the magnitude of an intervention program. In a small community-based program, key stakeholders may include the program director, staff, and clients. Stakeholder groups of a large federal program, on the other hand, could include federal agencies, state agencies, community-based organizations, university researchers, clients, program directors, program administrators, implementers, community advocates, computer experts, and so on.

Evaluators are usually hired by decision makers, and one of the major purposes of program evaluation is to provide information to decision makers that they will use to allocate funds or determine program activities. This contractual arrangement has a potential to bias evaluators toward the groups in power, that is, the decision makers who hire them or the stakeholders with whom the decision makers are most concerned. Critics such as House (1980) argued that evaluation should address social justice and specifically the needs and interests of the poor and powerless. However, Scriven (1997) and Chelimsky (1997) were concerned that when evaluators take on the role of program advocates, their evaluations' credibility will be tarnished.

Social justice is a difficult issue in evaluation. Participatory evaluation has the potential to alleviate some of the tension between serving social justice and decision makers. Including representatives of the various stakeholder groups in evaluation has been proposed as a way to address some social justice issues. Generally, stakeholders participate in an evaluation for two purposes: practical and transformative (Greene, Lincoln, Mathison, Mertens, & Ryan, 1998). Practical participatory evaluation is meant to enhance evaluation relevance, ownership, and utilization. Transformative participatory evaluation seeks to empower community groups to democratize social change. Either way, participatory evaluation can provide evaluators with an opportunity to engage with different stakeholder groups and balance diverse views, increase buy-in from all stakeholder groups, and enhance their willingness to use evaluation results.

Another way of enhancing evaluators' credibility is to promote professional ethics. Like other professionals, evaluators must adhere to professional ethics and standards. The American Evaluation Association (2004) adopted the following ethical principles for evaluators to follow:

- *Systematic inquiry.* Evaluators conduct systematic, data-based inquiries.
- *Competence.* Evaluators provide competent performance to stakeholders.
- *Integrity/honesty.* Evaluators ensure honesty and integrity of the entire evaluation process.
- *Respect for people.* Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders.
- *Responsibilities for general and public welfare.* Evaluators articulate and take into account the diversity and values that may be related to the general and public welfare. (“The Principles”)

In addition, to ensure the credibility of evaluation, the Joint Committee on Standards for Education (Yarbrough, Shulha, Hopson, & Caruthers, 2011) has specified the following five core standards for evaluators to follow:

1. *Utility standards.* The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.
2. *Feasibility standards.* The feasibility standards are intended to increase evaluation effectiveness and efficiency.
3. *Propriety standards.* The propriety standards support what is proper, fair, legal, right, and just in evaluations.
4. *Accuracy standards.* The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.
5. *Evaluation accountability standards.* The evaluation accountability standards encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement of and accountability for evaluation processes and products.

EVALUATION STEPS

The Centers for Disease Control and Prevention (CDC) published the *CDC Framework of Program Evaluation for Public Health* (CDC, 1999) to help evaluators understand how to conduct evaluation based on evaluation standards. The document specified six steps that are useful guides to the evaluation of public health and social betterment programs:

Step 1: Engage Stakeholders deals with engaging individuals and organizations with an interest in the program in the evaluation process.

Step 2: Describe the Program involves defining the problem, formulating program goals and objectives, and developing a logic model showing how the program is supposed to work.

Step 3: Focus the Evaluation Design determines the type of evaluation to implement, identifies the sources needed to implement the evaluation, and develops evaluation questions.

Step 4: Gather Credible Evidence identifies how to answer the evaluation questions and develop an evaluation plan that will include, among other things, indicators, data sources and methods for collecting data, and the timeline.

Step 5: Justify Conclusions involves collecting, analyzing, and interpreting the evaluation data.

Step 6: Ensure Use and Share Lessons Learned identifies effective methods for sharing and using the evaluation results.

EVALUATION DESIGN AND ITS COMPONENTS

When proposing an evaluation to stakeholders or organizations such as funding agencies, evaluators must describe the evaluation's purposes and methodology. An evaluation design needs to include at least five components:

1. *Purposes of and Background Information about the Intervention Program.* The first thing that evaluators need to do when assessing an intervention program is to gain a solid knowledge of the background of the program and document this understanding. Background information includes the purposes of the intervention program, the target population, the organizations responsible for implementing the program, key stakeholders of the program, implementation procedures, reasons for conducting the evaluation, the evaluation's timeline, the resources that will be used, and who will utilize the evaluation results. Evaluators usually gather information by reviewing existing documents such as program reports and the grant application proposal, as well as by interviewing key stakeholders of the program. The background information serves as a preliminary basis for communication by evaluators and stakeholders about the program and evaluation.

2. *A Logic Model or Program Theory for Describing the Program.* A sound evaluation requires a systematic and coherent description of the intervention program, which will serve as a basis for communication between evaluators and stakeholders and for the evaluation design. In reality, a systematic and coherent program description is often not available. It is unwise for evaluators to conduct a program evaluation without a mutual agreement with stakeholders about what the program looks like. In this situation, how could an evaluation provide useful information to stakeholders? Or, even worse, stakeholders later could easily claim that an evaluation failed to accomplish what they expected from it, if the evaluation results do not convey good news. Program description is an important step in evaluation.

If a program does not have a systematic and coherent program description, evaluators must facilitate stakeholders in developing one. This book discusses two options for describing a program: logic models and program theory. Logic models are used to identify the major components of a program in terms of a set of categories such as inputs, activities, outputs, and outcomes. However, if evaluators and stakeholders are interested in looking into issues such as contextual factors and causal mechanisms, this book encourages the use of program theory. Both logic models and program theory will be discussed in Chapter 3.

3. *Assertion of a Program's Stage of Development.* As will be discussed in the next chapter, an intervention program's life cycle can be generally classified as being in one of four phases: planning, initial implementation, mature implementation, and outcome. Program designers, during the planning phase, work with partners to identify or develop an intervention and organize resources and activities for supporting the intervention. After the planning phase, the program goes into the initial implementation phase. The major tasks here are training implementers, checking clients' acceptance, and ensuring appropriate implementation. After the initial implementation, the program progresses to the mature implementation stage. The major tasks here include ensuring or maintaining the quality of implementation. During the outcome phase, the program is expected to have desirable impacts on clients. The different stages of a program require different evaluation approaches. For example, constructive evaluation is most useful to a program during the initial implementation stage when it can help with service delivery, but it is not appropriate for a formal assessment of a program's merits at the outcome stage.

Evaluators and stakeholders have to agree on which stage a program is in to select an appropriate evaluation type(s) and approach. Chapter 2 will provide detailed discussions of the nature of program stages and how they relate to different evaluation types and approaches.

4. *Evaluation Types, Approaches, and Methodology.* This component is the core of evaluation design. Using information regarding the evaluation's purposes and the logic model/program theory, evaluators and stakeholders need to determine what type of evaluation, whether one of the basic evaluation types—constructive process, conclusive process, constructive outcome, or conclusive outcome—or a hybrid type, is suitable for correctly evaluating the program. Once program stage and evaluation type are determined, evaluators can move on to select or design an evaluation approach or approaches for evaluating a program. Chapter 2 will provide a comprehensive typology for guiding evaluators in selection of evaluation types and approaches.

Determining the most appropriate evaluation approach is challenging and time-consuming. However, it ensures that all involved share a mutual understanding of why a particular evaluation type has been selected. Without it, stakeholders are likely to find that the results of the evaluation address issues that are not of concern to them and/or are not useful to them. Stakeholders are often not trained on evaluation techniques. They often do not express what they expect and need from an evaluation as clearly and precisely as evaluators could hope. Evaluators usually must double- or even triple-check with stakeholders to make sure everyone shares the same understanding and agrees on the evaluation's purposes up front.

5. *Budget and Timeline.* Regardless of stakeholders' and evaluators' visions of an ideal evaluation plan, the final evaluation design is bound to be shaped by the money and time allocated. For example, if stakeholders are interested in a rigorous assessment of an intervention program's outcomes but can provide only a small evaluation budget, the research method used in the evaluation is not likely to be a randomized controlled trial over a few years, which would likely cost over a few million dollars. Similarly, if the timeline is short, evaluators will likely use research methods such as rapid assessments rather than conduct a thorough evaluation.

When facilitating stakeholders in making an informed decision, it is highly preferable for evaluators to propose a few options and explain the information each option is likely to provide, as well as the price tag of each.

MAJOR CHALLENGES OF EVALUATION: LESSONS LEARNED FROM PAST PRACTICE

Program evaluation has been practiced over several decades. Lessons learned from experience indicate that program evaluation faces a set of unique challenges that are not faced by other disciplines.

Judge a Program Not Only by Its Results but Also by Its Context

One important characteristic distinguishing program evaluation is its need, rarely shared by other disciplines, to use a holistic approach to assessment. The holistic approach includes contextual or transformation information when assessing the merit of a program. By comparison, product evaluation is more streamlined, perhaps focusing solely on the intrinsic value of its object. Products like televisions can be assessed according to their picture, sound, durability, price, and so on. In many situations, however, the value of a program may be contextual as well as intrinsic or inherent. That is, to adequately assess the merit of a program, both its intrinsic value and the context in which that value is assigned must be considered together. For example, say an educational program has, according to strictly performance-based evaluation, attained its goals (which are its intrinsic values). But in what context was the performance achieved? Perhaps the goal of higher student scores on standardized tests was attained by just “teaching students the tests.” Does the program’s performance still deserve loud applause? Probably not.

Similarly, what about a case in which program success is due to the participation of a group of highly talented, well-paid teachers with ample resources and strong administrative support, but the evaluated program is intended for use in ordinary public schools? This “successful” program may not even be relevant, from the viewpoint of the public schools, and is not likely to solve any of their problems. Therefore, *how* a program achieved its goals is just as important as *whether* it achieved them. For example, an outcome evaluation of one family-planning program in a developing country limited its focus to the relationship between program inputs and outputs; it appeared possible, on this basis, to claim success for the program. A large drop in the fertility rate was indeed observed following the intervention. Transformation information, however, showed that such a claim was misleading. Although the drop in fertility was real, it had little to do with the intervention. A larger factor was that, following implementation, a local governor of the country, seeking to impress his prime minister with the success of the program, ordered soldiers to seize men on the streets and take them to be sterilized. An evaluator with a less holistic approach might have declared that the goals of the program were attained, whereas other people’s personal knowledge led them to condemn the program as inhumane. Lacking a holistic orientation, program evaluation may reach very misleading conclusions.

Evaluations Must Address Both Scientific and Stakeholder Credibility

Program evaluation is both a science and an art. Evaluators need to be capable of addressing both scientific and stakeholder credibility in an evaluation. The scientific credibility of program evaluation reflects the extent to which that evaluation was governed by scientific principles. Typically, in scientific research, scientific credibility is all that matters. The more closely research is guided by scientific principles, the greater its credibility. However, as an applied science, program evaluation also exhibits varying degrees of *stakeholder* credibility. The stakeholder credibility of a program evaluation reflects the extent to which stakeholders believe the evaluation's design gives serious consideration to their views, concerns, and needs.

The ideal evaluation achieves both high scientific and high stakeholder credibility, and the two do not automatically go hand in hand. An evaluation can have high scientific credibility but little stakeholder credibility, as when evaluators follow all the scientific principles but set the focus and criteria of evaluation without considering stakeholders' views and concerns. Their evaluation will likely be dismissed by stakeholders, despite its scientific credibility, because it fails to reflect the stakeholders' intentions and needs. For example, there are good reasons for African-Americans to be skeptical of scientific experiments that lack community input, due to incidents such as the Tuskegee syphilis experiment (Jones, 1981/1993). Researchers in the experiment withheld effective treatment from African-American men suffering from syphilis so that the long-term effects of the disease could be documented. Conversely, an evaluation overwhelmed by the influence of stakeholders, such as program managers and implementers, may neglect its scientific credibility, resulting in suspect information.

One of the major challenges in evaluation is how to address the tension between scientific credibility and stakeholder credibility. Evaluation theorists, such as Scriven (1997), argued that objectivity is essential in evaluation because without it, evaluation has no credibility. On the other hand, Stake (1975) and Guba and Lincoln (1981) argued that evaluations must respond to stakeholders' views and needs in order to be useful. Both sides make good points, but objectivity and responsiveness are conflicting values. How would evaluators address this tension?

One strategy is to prioritize, choosing one type of credibility to focus on. However, this prioritization strategy does not satisfactorily address the conflict between the two values. A better strategy, proposed by and used in this book, is perhaps to strike a balance between the two. For example, evaluators might

pursue stakeholder credibility in the earliest phases of evaluation design but turn their attention toward scientific credibility later in the process. Initially, evaluators experience a great deal of interaction and communication with a program's stakeholders for the specific purpose of understanding their views, concerns, and needs. Evaluators then incorporate the understanding they have acquired into the research focus, questions, and design, along with the necessary scientific principles. From this point on, to establish scientific credibility, the evaluators require autonomy to design and conduct evaluations without interference from stakeholders. Stakeholders are usually receptive to this strategy, especially when evaluators explain the procedure to them at the beginning of the process. While stakeholders do not object to a program being evaluated, or dispute the evaluator's need to follow scientific procedures, they do expect the evaluation to be fair, relevant, and useful (Chen, 2001).

As will be discussed in the rest of the book, the tension between scientific and stakeholder credibility arises in many situations. Such tension makes evaluation challenging, but resolving it is essential for advancing program evaluation.

Evaluations Must Provide Information That Helps Stakeholders Do Better

Earlier in this chapter, we learned that Scriven placed a higher priority on conclusive assessment than on program improvement, while Cronbach preferred otherwise. This is an important, but complicated, issue for evaluators. Many evaluators quickly learn that stakeholders are eager to figure out what to do next in order to make a program work better. Stakeholders find evaluations useful if they both offer conclusions about how well programs have worked *and* provide information that assists the stakeholders in figuring out what must be done next to maintain—or even surpass—program goals. Thus, the assessment of a program's performance or merit is only one part of program evaluation (or, alone, provides a very limited type of evaluation). To be most useful, program evaluation needs to equip stakeholders with knowledge of the program elements that are working well and those that are not. Program evaluation in general should facilitate stakeholders' search for appropriate actions to take in addressing problems and improving programs. There are important reasons why evaluations must move beyond narrow merit assessment into the determination of needed improvements. In the business world, information on product improvement is provided by engineering and market

research; likewise, in the world of intervention programs, the agency or organization overseeing an effort relies on program evaluation to help it continually guarantee or improve the quality of services provided.

Consider that intervention programs typically operate in the public sector. In the private sector, the existence or continuation of a product is usually determined by market mechanisms. That is, through competition for consumers, a good product survives, and a bad product is forced from the market. However, the great majority of intervention programs do not encounter any market competition (Chen, 1990). Drug abusers in a community may find, for example, that only one treatment program is available to them. In the absence of an alternative, the treatment program is likely to continue whether or not its outcomes justify its existence. Furthermore, well-known programs with good intentions, such as Head Start, would not be discontinued based on an evaluation saying the programs were ineffectual; decision makers rarely use program evaluation results alone to decide whether a program will go on.

Under these circumstances, an evaluation that simply assesses the merit of a program's past performance and cannot provide stakeholders with insights to help them take the next step is of limited value (Cronbach, 1982). In fact, many stakeholders look to a broad form of program evaluation to point out apparent problems, as well as strengths upon which to build. In general, to be responsive and useful to stakeholders, program evaluation should meet both assessment needs *and* improvement needs rather than confine itself solely to conclusive assessment. Stakeholders need to know whether the program is reaching the target group, the treatment/intervention is being implemented as directed, the staff is providing adequate services, the clients are making a commitment to the program, and the environment seems to be helping the delivery of services. Any part of this information can be difficult for stakeholders to collect; thus, program evaluators must have the necessary training and skills to gather and synthesize it all systematically.

In a broad sense, therefore, merit assessment is a means, rather than the end, of program evaluation. Our vision of program evaluation should extend beyond the design of supremely rigorous and sophisticated assessments. It is important to grasp that evaluation's ultimate task is *to produce useful information that can enhance the knowledge and technology we employ to solve social problems and improve the quality of our lives.*

Furthermore, as discussed in the last section, constructive evaluation for program improvement and conclusive evaluation for merit assessment are not mutually exclusive categories. Evaluation does not have to focus on either program improvement or merit assessment. The introduction of hybrid evaluation types in this book provides options by which evaluation can address both issues.

ADDRESSING THE CHALLENGES: THEORY-DRIVEN EVALUATION AND THE INTEGRATED EVALUATION PERSPECTIVE

To better address these challenges, this book applies the frameworks provided by the theory-driven evaluation approach and the integrated evaluation perspective.

Theory-Driven Evaluation Approach

The theory-driven evaluation approach requires evaluators to understand assumptions made by stakeholders (called program theory) when they develop and implement an intervention program. Based on stakeholders' program theory, evaluators design an evaluation that systematically examines how these assumptions operate in the real world. By doing so, they ensure that the evaluation addresses issues in which the stakeholders are interested. The usefulness of the theory-driven evaluation approach has been discussed intensively in the evaluation literature (e.g., Chen, 1990, 2005, 2012a, 2012b; Chen & Rossi, 1980, 1983a; Chen & Turner, 2012; Coryn, Noakes, Westine, & Schröter, 2011; Donaldson, 2007; Funnell & Rogers, 2011; Nkwake, 2013; Rossi, Lipsey, & Freeman, 2004; Weiss, 1998). The concept and application of program theory will be intricately discussed in Chapter 3.

It is important to know that theory-driven evaluation provides a sharp contrast to traditional method-driven evaluation. Method-driven evaluation views evaluation as mainly an atheoretical activity. Evaluation is carried out by following research steps of a chosen research method such as randomized experiments, survey, case study, focus group, and so on. Within this tradition, evaluation does not need any theory. If evaluators are familiar with the research steps of a particular method, then they can apply the same research steps and principles across different types of programs in different settings. To some degree, method-driven evaluation simplifies evaluation tasks. However, because the focus of method-driven evaluation is mainly on methodological issues, it often does not capably address stakeholders' views and needs. The theory-driven evaluation approach argues that while research methods are important elements of an evaluation, evaluation should not be dictated or driven by one particular method.

Because theory-driven evaluation uses program theory as a conceptual framework for assessing program effectiveness, it provides information not only on whether an intervention is effective but also how and why a program

is effective. In other words, it is capable of addressing the challenge discussed in the last section: The success of a program has to be judged not only by its results but also by its context. This approach is also useful for addressing the following challenge: Evaluation must be capable of providing information for stakeholders to do better. The theory-driven evaluation approach will be intensively discussed in Chapters 3, 7, 12, 13, and 14.

Integrated Evaluation Perspective

Program evaluation is challenging because it has to provide evaluative evidence for a program that meets two requirements. The first requirement is that the evaluative evidence must be credible; that is, program evaluation has to generate enough credible evidence to gain a scientific reputation. This requirement is called the scientific requirement. The second requirement is that the evidence must respond to the stakeholders' views, needs, and practices so as to be useful. Stakeholders are consumers of evaluation. Program evaluation has little reason to exist unless it is able to adequately serve stakeholders' needs. This requirement is called the stakeholder requirement.

Ideally, evaluations should meet both requirements, but in reality evaluators often find it difficult to meet both. On the one hand, they must apply rigorous methods to produce credible evidence. On the other hand, evaluators often find it difficult to apply rigorous methods—such as randomized controlled trials (RCTs)—to evaluate real-world programs given insufficient resources and short time lines. In many situations, administrative hindrances and ethnic concerns add barriers to such an application. Furthermore, even should these barriers be removed and a rigorous method applied, stakeholders may feel that the focus of the evaluation is then too narrow or too academic to be relevant or useful to them. The reason for this disconnect is that the stakeholders' views on community problems and how to solve them are quite different from the conventional scientific methods' underlying philosophy—reductionism. Reductionism postulates that a program is stable and can be analytically reduced to a few core elements. If a program can be reduced to core components, such as intervention and outcome, then an adjustment can be implemented and desirable changes will follow. Given this view, the evaluators' main task is to rigorously assess whether the change produces predetermined outcomes.

However, stakeholders' views on and experiences with social problems and addressing them in a community are more dynamic and complicated

than those assumed by reductionism. Their views can be characterized as the following:

1. *An intervention program is implemented as a social system.* In a social system, contextual factors in a community—such as culture, norms, social support, economic conditions, and characteristics of implementers and clients—are likely to influence program outcomes. As discussed at the beginning of this chapter, program interventions are open systems, not closed like a biological system in terms of contextual factors.

2. *Health promotion/social betterment programs require clients, with the help of implementers, to change their values and habits in order to work.* Unfortunately, people are notoriously resistant to changing their values and habits. For example, an education program may require children fond of playing video games to substantially cut down on game playing to make time for studying; these children may vastly prefer playing the latest zombie massacre game to studying. Victims of bullying in schools may be asked to start reporting bullying incidents to school authorities and parents; based on past experience, these victims may believe reporting these incidents is useless or even dangerous. Because an intervention requires changes, its demands may be highly challenging to both clients and implementers. Not only must program designers wrestle with this challenge when designing an effective intervention program but evaluators must also take this reality into consideration when designing a useful evaluation.

Because of the above factors, stakeholders believe that they need to take a much broader approach in solving a community problem. An intervention is not a stand-alone entity but, rather, has to connect to contextual factors and/or change clients' values and habits to work. Their broad view of community problem solving is inconsistent with the traditional scientific methods, which focus on narrow issues such as assessing the causal relationships between an intervention and its outcomes. The inconsistency between stakeholders' views and reductionism's assumptions regarding community problems and interventions is partly why there is such a huge chasm between the academic and practice communities regarding interventions, as will be discussed in Chapter 15.

Stakeholders respect the value and reputation of scientific methods but view the information provided by using them as just one piece of a jigsaw puzzle they need to assemble. They need other pieces to complete the picture. They hope evaluators can figure out ways to provide all, not just one, of those pieces to them. Stakeholders are concerned that, if evaluators focus too much

on the scientific piece, it will blind them or prevent them from simultaneously investigating other means to solve the puzzle. Stakeholders' views on community problem solving are relevant to ideas proposed by systems thinking (e.g., Meadows, 2008). According to systems thinking, a system is made up of diverse and interactive elements and must address environmental turbulence. Problem solving thus requires the modification of groups of variables simultaneously.

The above analysis shows that evaluators face a dilemma in meeting the scientific requirement and the responsiveness requirement at the same time. An evaluation emphasizing the scientific requirement may scarify the responsiveness requirement, and vice versa. The dilemma has significant implications for evaluation practices, but it has not been intensively and systematically discussed in the literature. There are three general strategies evaluators use to address the dilemma:

Prioritizing the Scientific Requirement as the Top Priority in Evaluation: The first strategy is to stress the scientific requirement by arguing that evaluation's utility relies on whether it can produce credible evidence. Following this general strategy, evaluators must apply rigorous methods as best as they can. Issues related to the responsiveness requirement are addressed only when they do not compromise the rigor issues. Currently, this strategy is the most popular one used by evaluators (Chen, Donaldson, & Mark, 2011). The strategy appeals particularly to evaluators who are strongly committed to scientific values and evidence-based interventions.

Prioritizing the Responsiveness Requirement as the Top Priority in Evaluation. The second strategy is to put the emphasis on the responsiveness requirement. This strategy requires that evaluators use a participatory evaluation approach and qualitative methods to meet stakeholders' information needs (e.g., Cronback, 1982; Stake, 1975). This method is attractive to evaluators who view traditional scientific methods as too narrow and rigid to accommodate stakeholders' views and to meet their informational needs.

Synthesizing the Scientific and Responsiveness Requirements in Evaluation. The third general strategy is to synthesize the scientific and responsiveness requirements in evaluation. This strategy does not prioritize either requirement as the prime focus and thus avoids maximizing one at the expense of the other. Evaluations following this strategy may not be able to provide highly rigorous evidence but can provide good-enough evidence to balance the scientific and responsiveness requirements.

The first two strategies have merits. They are especially useful when there is a strong mandate for evaluation to be either highly rigorous or highly responsive. However, the author believes that, in many typical intervention programs, stakeholders are more likely to benefit from evaluations that use the synthesizing strategy. This book advocates this strategy and formally calls it the *integrated evaluation perspective*. Specifically, the integrated evaluation perspective urges evaluators to develop evaluation theories and approaches that can synthetically integrate *stakeholders' views and practices*, thus acknowledging the *dynamic* nature of an intervention program in a community, with *scientific principles and methods* for enhancing the usefulness of evaluation.

In spite of its conceptual appeals, the integrated evaluation perspective faces a challenge in developing specific evaluation theories and approaches to guide the work. It does not have advantages such as the scientific prioritization strategy. For example, advocates of the scientific prioritization strategy can borrow scientific methods and models developed by more matured disciplines and apply them to evaluation. The integrated evaluation perspective, however, does not have this ability because other disciplines do not face the kind of inconsistency between scientific and responsiveness requirements experienced in evaluation. They thus do not need to deal with synthesizing issues. For example, in biomedical research, both researchers and physicians consistently demand rigorous evidence for a medicine's efficacy. Accordingly, biomedical research cannot offer evaluators clues or solutions on synthesizing the conflict between scientific and responsiveness requirements. The integrated evaluation perspective, therefore, requires evaluators to develop innovative, *indigenous* theories and approaches to synthesize the requirements unique to the discipline.

This book contributes to the integrated evaluation perspective by introducing many innovative, indigenous theories and approaches evaluators can use in balancing the scientific and responsiveness requirements. At the same time, this book does not neglect traditional theories and approaches promoted by the scientific prioritization or responsiveness prioritization strategies. Instead, the author intends to introduce both traditional and innovative evaluation theories and approaches from these three strategies to enrich evaluators' toolbox so they can apply all theories and approaches as needed.

The nature and applications of the integrated evaluation perspective will be illustrated in detail in Chapters 11, 12, 13, 14, and 15, but its spirit and the principles it employs to develop indigenous concepts, theories, approaches, and methodologies are manifested throughout the book.

PROGRAM COMPLEXITY AND EVALUATION THEORIES

The discussion above of the dynamic and complicated nature of an intervention program in a community raises an interesting issue about program complexity. Evaluation theorists have different perceptions of how complex (e.g., in content, context, transformation, and stability) intervention programs are in general. Some may view these elements as quite stable or as fixed goals to achieve, whereas others view them as highly complex or fluid. How theorists view the complexities of a program can influence the theories or approaches they propose to use (Chen, 2012a).

To allow us to understand the issue, envisioning a continuum of program complexity is helpful, with reductionism at one end and fluid complexity at the other end. As discussed in the last section, reductionism postulates that a program is stable and can be analytically reduced to a few core elements.

On the other hand, fluid complexity, a concept created by the author for facilitating discussions, represents the view that an intervention program needs to constantly change its diverse and interactive elements to address ongoing environmental turbulences. Under the fluid complexity view, evaluators must speedily collect and analyze any available information on changes and promptly report the findings to decision makers to quickly adjust and readjust courses of action. For example, Christopher Columbus's expedition team not only had to constantly revise its plans for addressing ongoing external threats but also had to completely change its mission. Upon replacing the original mission of finding a route to India with the new mission of discovering a new world, the expedition was judged an enormous success.

Reductionism has its strengths and limitations. Reductionism has merits in its easy coexistence with known quantitative methodological and statistical models. Evaluators can use these methods and techniques to provide rigorous evidence of an intervention's efficacy or effectiveness. However, in its purest form, reductionism oversimplifies a program and provides an unsustainable solution. Fluid complexity also has its strengths and limitations. It may provide creative or sustainable solutions for complicated problems; however, at least for now, common quantitative methods and statistical models are not capable of effectively analyzing complex, fluid interactions. Furthermore, if a program is extremely complex and dynamic, then it lacks an entity for meaningful evaluation. Consultants are then more suitable than evaluators for offering opinions on how to assess and address problems.

Evaluation theories can be placed somewhere along the reductionist–fluid complexity continuum. Some are closer to reductionism; some, to fluid complexity. For example, the experimentation evaluation approach, which will be

discussed in Chapter 10, is closer to the reductionist end of the continuum. On the other hand, the developmental evaluation approach (Patton, 2011) is closer to the fluid complexity end of the continuum. The integrated evaluation perspective with its related theory-driven evaluation approach proposed in this book lies close to the middle. The perspective attempts to provide a synthesis of the different views proposed by reductionism and fluid complexity. It agrees with fluid complexity that the environment can create uncertainties and pressure stakeholders and evaluators to make changes, but it also proposes that a program can find proactive measures to reduce uncertainty and maintain a level of stability. For example, program managers and staff can build partnerships to buffer political pressure, and particular strategies, such as environmental scanning or problem-solving networks, can be helpful in reducing uncertainties. In applying the integrated evaluation perspective, this book proposes and examines many evaluation approaches and methods that consider variables and factors beyond what reductionism would recognize, but they are not too complex to be analyzed within existing quantitative and qualitative methods.

It is not clear where real-world programs fall along the continuum. However, it is clear that every theorist wishes to see or argue that the distribution congregates where his or her theory lies on the continuum. For example, this author would argue that real-world programs likely fall along the continuum in a normal (bell shape) distribution, with the majority at the middle. If this distribution proves true, the evaluation concepts, approaches, and methods proposed by this book are likely to be applicable to the majority of typical programs. Of course, other theorists would disagree. I encourage readers to form their own opinions and join the discussion.

WHO SHOULD READ THIS BOOK AND HOW THEY SHOULD USE IT

This book introduces practical evaluation approaches and methods to evaluators, but it avoids becoming a “cookbook.” The approaches and methods discussed here are supported by a context of underlying principles and theoretical justification. This context, it is hoped, delivers knowledge with the latitude and flexibility program evaluators need to design suitable evaluation models. With this in mind, this book was prepared for two audiences.

Students

The first anticipated audience is students, especially those interested in issues related to the practice of evaluation, including the challenges evaluators

can expect and practical means of dealing with them. The book may liberate such students from the notion that evaluations are mainly methodological activities. Students should not feel like mindless number crunchers. The book will challenge students to seek strategies for broadening basic social science theories learned in the classroom, linking these to action and intervention theories employed in the field by program staff, evaluators, and social reformers. This text is ideal for use as a textbook for the following evaluation courses:

Introductory evaluation course. For an introductory course, the author would recommend covering Chapters 1 to 12.

Advanced evaluation course. Because of its depth and comprehensive scope, the book can also be used as one of the books in an advanced evaluation course. In such a course, instructors are encouraged to cover all the chapters from Chapters 1 to 15.

Evaluation Practitioners

The second audience is evaluation practitioners, especially those who seek new knowledge to strengthen their practical skills or expand the scope of their work. Such practitioners should generally look to the book to broaden their vision of evaluation alternatives so as to increase their skill at designing evaluations that fit a variety of program circumstances and evaluation purposes. Seasoned program evaluators may find both valuable insights into established evaluation strategies and approaches and new, innovative ideas for further enhancing their practice.

INTRODUCING THE REST OF THE CHAPTERS

In Chapter 2, based upon the fundamental typology of evaluation types discussed in this chapter, a road map of evaluation options—the “comprehensive evaluation typology”—is presented. The typology can guide evaluators and stakeholders in selecting the approaches and methods best suited to meet a program’s circumstances and the stakeholders’ needs at different program stages (program planning, initial implementation, mature implementation, and outcome), as discussed in Chapters 4 through 12. Chapter 3 discusses logic models and program theory, which are the foundation for understanding and

describing a program as discussed throughout the book. Chapters 13 to 15 discuss cutting-edge issues in program evaluation.

As will be discussed in Chapter 2, this book can be applied to start-up programs or established programs. For a start-up program, evaluators may be asked to evaluate one or more program stages, choosing among the planning, initial implementation, mature implementation, and outcome stages. For an established program, evaluators typically are invited to conduct evaluation activities at the mature implementation stage and/or the outcome stage.

QUESTIONS FOR REFLECTION

1. Detail a real-world intervention program and discuss its inputs, transformation, outputs, and environment.
2. Why is the feedback stage necessary to the success of an intervention program?
3. Define formative and summative evaluations. Give examples of each type.
4. Give examples of constructive process, conclusive process, constructive outcome, and conclusive outcome evaluation types.
5. Compare and contrast the dual formative/summative distinction with the fundamental evaluation typology.
6. What are hybrid evaluations? Give examples of this type of evaluation.
7. As the head of an agency or organization, how would you ensure that an internal evaluator provides useful information? How would you ensure that an external evaluator provides useful information?
8. Why are politics so important when planning to conduct an evaluation?
9. List some examples of potential stakeholders in an intervention program and explain why evaluators need to engage them when designing and conducting an evaluation.
10. Explain why the success of a program cannot be judged only by its results. Give examples.
11. Explain why research may be able to focus mainly on scientific credibility, while evaluation must have both scientific and stakeholder credibility.

12. Why do evaluators face a dilemma in addressing scientific and responsiveness requirements? What are three strategies to address the dilemma? Which strategy would you take? Why?
13. What is the integrated evaluation perspective? What are the challenges faced by this perspective?
14. The author argued that the distribution of real-world programs along the program complexity continuum may be like a normal bell-shaped distribution. He also mentioned that other theorists would disagree with him. Why? In your view, what would the distribution look like? Provide examples and justifications supporting your argument.

CHAPTER 2



Understand Approaches to Evaluation and Select Ones That Work

The Comprehensive Evaluation Typology

This book will introduce a variety of evaluation approaches for readers to include in their toolbox. These approaches represent the “science” of program evaluation. However, simply having the tools does not make one a competent evaluator. An evaluator must also know how to find out stakeholders’ needs, explain various evaluation approaches to them and get their input, and select an appropriate approach to use. This process is the “art” in the art and science of program evaluation. A competent evaluator has mastered both the science and art of the discipline. However, there is much less discussion of the art than of the science. Much less is written about issues that affect whether the best approach is selected from among the many available, and the information that is available tends not to be systematically presented.

An analogy about fishing suggests how this gap in program evaluation can make life more difficult for the program evaluator. To go fishing one needs, first of all, equipment—the poles and lines, hooks, sinkers, floaters, and bait or lures. Without this basic equipment, fishing is (for most humans) impossible. Possessing equipment and knowing how to use it, however, do not guarantee success. Choosing the wrong equipment from one’s closet full of fishing tackle—the wrong size of fishing line, or the wrong hook, or an inappropriate bait for a given fishing spot—probably means ending the day empty-handed, even if one handles that line, hook, or bait magnificently.

The vital, yet limited, role of equipment in fishing is seen clearly in the art of fly-fishing, in which the angler continuously casts and retrieves a line tipped with, or baited with, an artificial fly. With tackle and casting know-how, any person can go through the motions of fly fishing. Only those who are accomplished in the art, however, can count on catching fish most days. A good fly-fishing angler knows how to choose the right place and time as well as the right artificial fly. Fly-fishing masters have learned to habitually consider such things as season, currents, play of light and shade, and surrounding vegetation in addition to their equipment. These masters can select just the right fly to mimic whatever real fly would inhabit a given area at a given moment so that the fish strike the fly without suspicion.

Productive fishing is more than equipment and the ability to operate it. To catch fish, it helps to know fish habits and habitat—favorite foods, favorite pools or banks, responses to weather, and so on. Upon this kind of understanding is based “the art of fishing.” Mastery of this art may be gained through trial and error over a long period of time or, more efficiently, from instruction by someone experienced in fishing. It may also be developed by studying authoritative books.

What can program evaluators learn from the art of fishing? To begin with, consider evaluation approaches alone as being analogous to fishing equipment. This is the scientific aspect of fishing. Only when we have our evaluation approaches down pat can we set about catching a fish. But although this is a necessary condition, it remains an insufficient one. Like the average fishing enthusiast with a tackle box, an evaluator familiar with evaluation approaches can try his or her luck, but with no more guarantee of producing a decent evaluation than of landing a catch. Like fishing in the wrong spot or using the wrong bait, missing important issues in evaluation design finds the adventurer returning home with no prize. Evaluation becomes productive only when we go beyond rote approaches to ply the waters with some knowledge of the art of evaluation. This knowledge tells us under what conditions stakeholders are more likely to have different evaluation needs, what kinds of evaluation approaches have the potential to meet their needs, how to describe these potential approaches to stakeholders, how to solicit their input, and so on. The comprehensive typology that includes this art is an expansion of the fundamental evaluation typology discussed in Chapter 1. This comprehensive typology provides a broader range of ideas about how to use the art of program evaluation to produce fruitful results.

THE COMPREHENSIVE EVALUATION TYPOLOGY: MEANS AND ENDS

The comprehensive evaluation typology is based on a program’s life cycle. An intervention program’s life cycle is characterized by the following four stages:

program planning, initial implementation, mature implementation, and outcome. Stakeholders' evaluation needs vary from one stage to another (Chen, 2005). Thus, evaluation approaches also differ. Program evaluators need to know which evaluation strategy and approach, out of the many available, will be best suited to meet stakeholders' needs at which program stage. The comprehensive typology of practical program evaluation shown in Table 2.1 provides evaluators with systematic guidance to weighing the circumstances and needs of a given evaluation assignment against the strengths and shortcomings of various evaluation strategies and approaches.

Table 2.1 A Comprehensive Evaluation Typology: Evaluation Purposes, Approaches, and Strategies by Program Stage

<i>Program Stages and Evaluation Purposes</i>	<i>Evaluation Approaches</i>	<i>Evaluation Strategies</i>
1. Program-Planning Stage Provide pertinent information and assistance to help stakeholders develop program rationale and plan.	<i>Constructive Tools</i>	
	Needs assessment Formative research	Background information provision
	Logic models Program theory	Development facilitation
	<i>Hybrid Tools</i>	
	Assumption testing Pilot-testing	Troubleshooting
	Bilateral empowerment evaluation	Partnership
	<i>Conclusive Tools</i>	
	Commentary or advisory meeting	Merit assessment
2. Initial Implementation Stage Provide timely information on implementation problems and resources that will assist stakeholders with fixing the problem and stabilizing the program.	<i>Constructive Evaluation</i>	
	Formative evaluation Program review/ development meeting	Troubleshooting
	Bilateral empowerment evaluation	Partnership

(Continued)

Table 2.1 (Continued)

<i>Program Stages and Evaluation Purposes</i>	<i>Evaluation Approaches</i>	<i>Evaluation Strategies</i>
3. Mature Implementation Stage Assess the quality of implementation to serve accountability needs. Monitor progress of implementation. Problem identification and solving Holistically assess implementation process.	<i>Conclusive Evaluation</i>	
	Process (fidelity) evaluation	Performance assessment
	Process monitoring	Performance monitoring
	<i>Constructive Evaluation</i>	
	Formative evaluation	Troubleshooting
	<i>Hybrid Evaluation</i>	
	Theory-driven process evaluation	Enlightenment assessment
4. Outcome Stage Improve the coherence of a program. Monitor the client's progress toward outcomes. Assess pure independent effects. Assess joint effects of a real-world program. Holistically assess program to serve accountability and program improvement needs. Assess transferability.	<i>Constructive Evaluation</i>	
	SMART goals	Development facilitation
	Evaluability assessment	Development facilitation
	Plausibility assessment/consensus building	Development facilitation
	<i>Conclusive Evaluation</i>	
	Outcome monitoring	Performance monitoring
	Validity-focused outcome evaluation	Performance assessment
	Viability evaluation	Performance assessment
	<i>Hybrid Evaluation</i>	
	Real-world outcome evaluation	Performance assessment Development facilitation
	Theory-driven outcome evaluation	Enlightenment assessment
	Transferability evaluation	Enlightenment assessment

Stages in the Program Life Cycle

The first row of Table 2.1 lays out the four stages of a program's development and highlights the evaluation requirements associated with each stage. Evaluators can best understand stakeholders' evaluation needs if informed of which stage(s) the stakeholders are interested in evaluating.

It is often assumed that a program will move sequentially through these stages. In reality, programs can go back and forth between stages in a nonlinear fashion. As an example, consider a program in its mature implementation stage. The program has been troubled by several major problems with service delivery, and its stakeholders decide to revise the program plan and return to the initial implementation stage. Thus, their evaluation needs are different from what they once were. Similarly, if in its outcome stage a program is found to be ineffective, its stakeholders could decide to redesign the program, returning to the initial planning stage.

Evaluators in the field are asked to conduct evaluations for programs at any stage and at various combinations of stages. When the program is established, evaluation of its implementation and outcome stages is common; start-up programs, too, need evaluations at these stages. Start-up programs also frequently require evaluation during the planning and initial implementation stages. The following paragraphs discuss the evaluation needs characteristic of each stage.

Program-Planning Stage

The first of the four stages is the program-planning stage. This is the very beginning. Stakeholders at this stage—for example, program designers—are developing a plan that will serve as a foundation for organizing and implementing a program at some future date. As we have seen, programs can be complex; stakeholders often seek considerable help from experts with the hope of developing a plan of high quality. Today, evaluators are often found among these experts. In the program-planning stage, stakeholders' primary evaluation need is to learn the evaluation concepts, strategies, and activities that can help in the design and development of a program rationale and a program plan.

Early in Part II of this book, we will read that program evaluation has, across much of its history, focused on outcomes. Lessons from the field, however, have plainly taught that program failures are often essentially *implementation* failures, and the focus of evaluation has gradually broadened to include processes as well as outcomes. The current view is that a major part of implementation

failure can be traced to poor program planning and development. Evaluators can make important contributions in these areas where attention is most needed.

Initial Implementation Stage

The second stage cited in Table 2.1 is the initial implementation stage. As a program plan begins to be put into action, much can go wrong. During the initial implementation stage, a program's course can be highly fluid and unstable. At this point, stakeholders need timely feedback on major implementation elements and identification of the sources of problems. These kinds of data can help stakeholders to troubleshoot implementation problems and quickly stabilize the program.

Mature Implementation Stage

The mature implementation stage begins when implementation of the program has settled into routine activities. Rules and procedures for conducting program activities are now well established. Stakeholders are likely to be interested in one or more of the following: continued unearthing of the sources of immediate problems, generation of data reassuring to those to whom stakeholders are accountable, and program improvement. Even in maturity, a program is subject to problems such as clients' dissatisfaction with services. A wise course for stakeholders in a case like this is to seek timely information from evaluators about the cause of problems. Identifying problems and resolving them are key to improving a program. Furthermore, as a program matures, stakeholders may think more about their accountability. Data illustrating the effectiveness of implementation or the efficiency of service delivery are useful to stakeholders, who often ask evaluators to find such data if they exist. Finally, within the mature implementation stage, stakeholders begin to look for strategies of improvement (tied to their need to be accountable, perhaps). They call on evaluators to provide holistic information through process evaluation with a purpose that goes beyond assessing the quality of implementation to strengthening the program processes.

Outcome Stage

The fourth stage of program development is known as the outcome stage. Following a period of program maturity, stakeholders inside and outside the program want to know whether the program is achieving its goals. An evaluation at this point can serve any of five primary evaluation needs. First, stakeholders may rely on evaluators to determine whether a program is ready for outcome evaluation.

It may not be, and, if it is not, evaluators may be asked for help in building the program's "evaluation capacity." Second, stakeholders may want to monitor their clients' progress. Third, stakeholders may ask for information on what the program would be achieving if it existed in the ideal environment. (Such information can also help stakeholders decide whether a program should be expanded to other people or settings.) Fourth, stakeholders may seek to know in detail the program's effects in its real-world setting because these, obviously, have a direct bearing on practice. Finally, some stakeholders may ask evaluators to go beyond traditional evaluation and its single-minded focus on assessment; they may want an evaluation that serves both accountability and program improvement needs.

Dynamics of Transition Across Program Stages

Intervention programs are goal-oriented activities. Ideally, the program moves directionally through the following stages: program-planning, initial implementation, mature implementation, and, finally, outcome. For the convenience of the reader, the remaining chapters of this book are arranged according to an ideal sequence of program stages. In reality, however, as noted above, program stages may not be linear at all. For example, due to a stakeholder's dissatisfaction with the direction of a program, or due to political pressure calling for a change, a program could move from the mature implementation stage back to the planning stage rather than moving forward to the outcome stage. Similarly, a program at the outcome stage may overhaul its operational procedures and move back to the initial implementation stage. This book accommodates the nonlinear transition of program stages and the evaluations related to each, as readers can pick and choose from the chapters and rearrange their reading about evaluation at each program stage to fit their program's development.

EVALUATION APPROACHES ASSOCIATED WITH EACH STAGE

An *evaluation approach* constitutes a systematic set of concrete procedures and principles that guide the design and conduct of an evaluation. The evaluation approach determines the evaluation's focus; it affects the research methods applied to collect and analyze data, as well as the interpretation of data. In Table 2.1, the second column lists the popular evaluation approaches associated with each stage. Evaluators can carry these tools in their toolbox and use them as needed. These approaches are classified in terms of the following functions: constructive evaluation, conclusive evaluation, or hybrid evaluation. Readers for

whom the terminology in the table is new should note that subsequent chapters of the book explore and explain the strategies and approaches in detail.

Planning Stage

Evaluation approaches in the planning stage can be classified into two categories: constructive and hybrid. These approaches are briefly described as follows.

Constructive Approaches

At the planning stage, evaluators could apply the following constructive tools to assist stakeholders with planning an intervention program:

Needs assessment. This approach is useful to identify service needs or gaps in a community. This tool is discussed in Chapter 4.

Formative research. This tool is used to conduct an in-depth inquiry to understand the nature and cause of a problem; it is discussed in Chapter 4.

Logic models and program theory. Evaluators use these tools to help stakeholders effectively describe their programs. A logic model or program theory is an essential foundation for program planning and evaluation; every evaluation should have this component. These tools are discussed in Chapter 3.

Hybrid Approaches

Relevancy testing. This tool is used to test how realistic a logic model or program theory is. It is discussed in Chapter 4.

Pilot testing. This tool, also discussed in Chapter 5, is used to test the feasibility of a program in the field.

Commentary or advisory meeting. Stakeholders and evaluators use this tool to solicit experts' opinion on the proposed program's overall merits or drawbacks and/or specific strengths and weaknesses for improvement; this tool is discussed in Chapter 5.

Initial Implementation

Constructive Evaluation

Formative evaluation. The purpose of this evaluation is to troubleshoot any problems in the early implementation stage. A full discussion of this evaluation is in Chapter 6.

Program review/development meeting. The purpose of this meeting is to ask staff about their experiences to identify any difficulties they are having with implementing the program. This evaluation is discussed in Chapter 6.

Bilateral empowerment evaluation. This partnership between evaluators and stakeholders facilitates a mutual learning process that can power the ongoing improvement of a program. This evaluation is discussed in Chapter 6.

Mature Implementation

Conclusive Evaluation

Process (fidelity) evaluation. The purpose of this evaluation, discussed in Chapter 7, is to assess whether an intervention has been implemented as it was intended.

Process monitoring. This evaluation monitors intervention implementation across months or years, providing general information on whether implementation is moving in an acceptable direction. It is discussed in Chapter 8.

Hybrid Evaluation

Theory-driven process evaluation. This evaluation not only assesses whether an implementation is appropriate or not, but also how and why it is or isn't working. This evaluation is discussed in Chapter 7.

Constructive Evaluation

Formative evaluation. Formative evaluation can still be applied at the mature stage. If stakeholders suspect some areas of a program may have problems, formative evaluation can be applied to identify and troubleshoot them.

Outcome Stage

Constructive Outcome Evaluation

The purpose of constructive outcome evaluation is to strengthen a program's coherence so as to enhance the program's success.

SMART goals. This tool is used to ensure that program goals are clear and measurable. This type of evaluation will be discussed in Chapter 9.

Evaluability assessment. This assessment, also discussed in Chapter 9, is done to ensure that a program can be evaluated.

Plausibility assessment/consensus building. This tool ensures that stakeholders view the program goals as feasible and agree on the goals. This tool is discussed in Chapter 9.

Conclusive Outcome Evaluation

Validity-focused outcome evaluation. This evaluation stresses rigor in assessing an intervention's efficacy or effectiveness. This evaluation is discussed in Chapter 10.

Outcome monitoring. This tool, discussed in Chapter 8, provides general information on whether a program is moving in a desirable direction.

Viability evaluation. The purpose of this evaluation is to assess whether an intervention is likely to survive in the real world; it is discussed in Chapter 15.

Hybrid Outcome Evaluation

Real-world outcome evaluation. This evaluation consists of both constructive and conclusive outcome assessment of a real-world program. It is discussed in Chapter 11.

Theory-driven outcome evaluation. This evaluation provides information not only about whether a program is effective, but also on how and why it is working. This evaluation is discussed in Chapters 12, 13, and 14.

Transferability evaluation. This evaluation provides information useful for assessing whether a program is transferable from one setting to another. It is discussed in Chapter 15.

STRATEGIES UNDERLYING EVALUATION APPROACHES

The above evaluation approaches are usually based upon some general strategies. An evaluation strategy is the general path that the evaluator and stakeholders take or orientation they have in order to fulfill a given evaluation approach's purpose. An understanding of these strategies prepares evaluators to better apply these approaches.

The last column of Table 2.1 lists a strategy used by the corresponding evaluation approach. For example, merit assessment is one general strategy employed to serve stakeholders' accountability needs. Other evaluation strategies include the development and enlightened strategies. The overall evaluation strategy must be closely related to the stakeholders' evaluation needs. As seen

in Table 2.1, typically several evaluation approaches are harnessed together within one evaluation strategy.

Whereas most stakeholders are unfamiliar with evaluation approaches or individual research procedures and techniques (evaluators are assumed to be familiar with these), they are usually acquainted with the general directions that evaluation strategies can take. For example, stakeholders may not know what a quasi-experiment is, but they do understand generally what a strategy such as merit assessment entails. The easiest and best course for evaluators is to determine the appropriate program evaluation approach by discussing with stakeholders the strategies the evaluators think will fit the stakeholders' evaluation needs. With stakeholders' input in mind, the evaluators can then lead a discussion of various appropriate evaluation approaches. Too often, evaluators neglect to engage in dialogue about evaluation strategies, launching right into the selection of evaluation approaches. The fallout from this practice can be stakeholders' *uninformed* consent to employ whatever evaluation approach the evaluator recommends. With little or no understanding of what that approach consists of, stakeholders may, when handed the final report, realize that it is not what they wanted and does not provide the information they need.

My view of evaluation strategies and approaches has been greatly influenced by my acquaintance with the following case of misdirected evaluation as briefly mentioned before. The client was a group of high-performing community-based organizations seeking to provide capacity-building services to similar but less accomplished organizations. A skilled and respected evaluator carried out the project. Before beginning, this evaluator met with stakeholders several times to discuss potential evaluation approaches. The parties decided to adopt mixed methods to assess the results of the capacity-building program. The final evaluation report provided a detailed pros-and-cons assessment of the program, expressing in general a favorable position toward the project. Unfortunately, those anticipated to be the program's service providers complained that the evaluation offered few insights into improving their program. The generated information was of the wrong kind, they protested, because the evaluation failed to reflect their needs and views. In the end, it became clear that the service providers had wanted a construction-oriented evaluation, whereas the evaluator had conducted a conclusion-oriented evaluation.

This is not a case easily dismissed by blaming the service providers for misstating their evaluation needs at the beginning or for changing their minds later on. Upon review of the project with the evaluator and the service providers, it appeared that the heart of this problem was an absence of effective tools with which stakeholders could voice their evaluation needs and identify suitable accompanying evaluation approaches.

This comprehensive evaluation typology illustrated in Table 2.1 includes both evaluation approaches that are frequently used and cutting-edge approaches that will be greatly used in the future. This chapter intends mainly to provide a bird's-eye view of various evaluation strategies and approaches that evaluators and stakeholders could choose at a particular program stage. Readers for whom the terminology in the table is new should note that subsequent chapters of the book explore and explain the strategies and approaches in detail.

To begin, there are four general categories of strategies included in the comprehensive typology of program evaluation: merit assessment, development, enlightenment, and partnership.

Merit Assessment Strategies

Merit assessment strategies are those that can provide information about the performance or merit of a program. Two merit assessment strategies frequently used by evaluators are the performance assessment strategy and the performance-monitoring strategy.

Performance Assessment Strategy

Performance assessment is the use of rigorous designs to provide credible information about a program's merit in terms of either its implementation process or its outcomes. The performance assessment strategy is part of a long-standing, influential tradition in program evaluation. As discussed in the section on evaluation types in Chapter 1, the performance assessment of a program does not have to wait till the outcome stage; it could happen even at the planning stage. For example, stakeholders could ask evaluators or independent experts to conduct a conclusive evaluation of whether a program plan should move forward or not. Performance assessment can also be carried out at the mature implementation stage. Fidelity evaluation is a type of process evaluation that assesses whether a program has been or is being implemented according to expectations. Another approach popularly used with the performance assessment strategy is the *outcome evaluation*, which assesses a program's success in reaching its goals.

Performance-Monitoring Strategy

The *performance-monitoring strategy* uses indicators to follow the implementation process and outcomes of a program across time. In a drug treatment program, for example, evaluators might monitor clients' drug use both before and after they experience the intervention. The performance-monitoring strategy comprises two well-known approaches: *process monitoring* and *outcome*

monitoring. Process monitoring cannot produce as much in-depth information about a program's implementation as does process evaluation; neither is outcome monitoring likely to produce convincing data about an intervention's effect on outcomes the way outcome evaluation can. In their defense, however, process monitoring and outcome monitoring are useful for managing a program and likely to cost less than typical process evaluation and outcome evaluation.

Development Strategies

Development strategies collect evaluative data relatively quickly in order to assist stakeholders with program planning or development. Three development strategies are well established in program evaluation: the background information provision strategy, the troubleshooting strategy, and the development facilitation strategy.

Background Information Provision Strategy

Evaluators use the *background information provision strategy* to research a program's community characteristics and needs, target population characteristics, and/or intervention options. The information gathered should help program designers and other stakeholders plan or strengthen a program. Evaluation approaches suited to this strategy include needs assessment and formative research. Needs assessment refers to determining and prioritizing the needs of a community or target population, such as when an agency asks what kinds of youth services are most needed in a community. In such a case, program evaluators might systematically interview youths, parents, and community leaders to help the agency answer its question. Formative research differs from needs assessment in that it places greater emphasis on identifying or prioritizing needs. Formative research consists of gathering empirical information on community and target population characteristics, as well as intervention options, to help stakeholders plan and develop programs. For example, program designers who are uncertain what kind of drug prevention program would be best received by new immigrants might engage program evaluators to manage a survey or focus group meeting, obtaining information that enables the program designers to make a decision.

Troubleshooting Strategy

The *troubleshooting strategy* is a system for identifying difficulties that programs are having and addressing them. The troubleshooting strategy is used,

first, to provide a timely assessment of barriers and/or problems facing a program; its second use is to identify options available to stakeholders to address difficulties. The value of this strategy lies in its potential ability to effectively identify an implementation problem before major damage occurs. Evaluators using this strategy must also provide stakeholders with information that facilitates resolution of the problem. The troubleshooting strategy is associated with use of the formative evaluation, relevancy testing, pilot-testing, and commentary and advisory approaches.

Formative evaluation is associated with research methods that are flexible to use, are easy to adopt in the field, and have a short turnaround time. For example, focus groups and participant observations can be used to collect, in timely fashion, facts about barriers and problems in implementation that then can be used to strengthen the program. Having chosen to target newly arrived immigrants, for instance, an HIV prevention program further decides to serve them with group counseling. After the implementation is carried out, evaluators are contracted to look for potential problems in the recently completed process. Using formative evaluation, evaluators interview a sample of the clients and quickly learn that some clients—Asian immigrants—are uncomfortable in group discussions of sexual behavior. The prompt feedback made available to the program director by the formative evaluation approach prompted modification of the program to better serve this particular immigrant group.

Formative evaluation and formative research (an approach affiliated with the background information provision strategy) are both research activities, yet with an important difference: Whereas formative evaluation examines directly the program's implementation, formative research is usually carried out before implementation and produces background information related to program planning. For example, the evaluator tackling the above assignment from a formative research approach might study a target group's cultural background as it relates to sexual behavior in hopes of facilitating program design decisions. An evaluator using the formative evaluation approach would evaluate the given target population's experience with the program itself.

The troubleshooting strategy also includes *reality testing*. Reality testing is the small-scale assessment of causal assumptions underlying a program and whether these assumptions hold up in the field. Reality testing can be used to strengthen the soundness of a change model. In contrast, the *pilot-testing approach* to the troubleshooting strategy involves actually operating the program on a very small scale. Unlike relevancy testing, pilot testing usually focuses on the action model. The information and experience gained from pilot testing can help strengthen a program before formal implementation begins

because areas needing modification can be fixed early and prevented from affecting the full-scale implementation. Another troubleshooting strategy is the *commentary and advisory approach*, in which data are not from the field. Instead, the expertise of evaluators is tapped as they review and comment on an existing action model and change model. They advise stakeholders about probable strengths and weaknesses of the model and offer suggestions for improvement. Finally, the *program review/development meeting approach* generates insights through systematic discussions in a meeting format among a group of program implementers and staff. With the evaluator providing facilitation, the experiences of a program are discussed, any implementation problems are dissected, barriers to and facilitators of these problems are identified, and strategies are developed to strengthen the program.

Development Facilitation Strategy

Evaluators' knowledge and skills are also central to the *development facilitation strategy*, which is defined in this book as the use of such expertise to help key stakeholders in a meeting/workshop setting. The development facilitation strategy functions to facilitate the stakeholders' efforts to develop or fine-tune the logic of a program or to identify its problems and seek programmatic solutions for them. Using this strategy, evaluators become facilitators and consultants, essentially; Patton (2011) and Guba and Lincoln (1989) have emphasized the value of this method for solidifying a common vision, winning support, and broadening a program's capacity. Expert evaluators can draw on their program evaluation skills to contribute greatly to the development of coherent programs that are logical in their foundations and feasible to implement.

Some evaluation approaches associated with the development facilitation strategy are the *conceptualization facilitation approach*, such as logic models and program theory discussed in Chapter 3. In the case of program theory, the conceptualization facilitation approach requires evaluators to work as facilitators and consultants, clarifying stakeholders' ideas about their program theory ideas, especially those concerning action and change models, and then facilitating their efforts to develop these models.

Enlightenment Strategy

Stakeholders may, of course, seek program evaluation in response to accountability needs as well as those of program improvement. As discussed in Chapter 1, program improvement remains the ultimate goal of program evaluation, and pure performance evaluation has little to say about improving programs.

Nonetheless, it is possible to design evaluations to meet both kinds of needs. The key is to extend the evaluation beyond conclusive assessment by examining the underlying assumptions and mechanisms that mediate the effects of the program. Evaluators with this orientation are practicing the *enlightenment strategy*. Enlightenment strategy is discussed at length in the literature. In general, it arises from the position that “assessment is means, program improvement is end.” The work by Cronbach (1982) is regarded as a pioneering in the development of the enlightenment strategy.

Enlightenment strategy is usually applied via a hybrid type of evaluation, that is, one that provides both conclusive and constructive information. For example, in the program-planning stage, evaluators could be asked to do a hybrid conclusive/constructive assessment of a program plan. They would provide information on the overall merit of the plan as well as detailed, concrete suggestions for how to improve the plan. Similarly, evaluators could conduct a conclusive/constructive process evaluation at the mature implementation program stage by not only assessing whether a program has been or is being implemented according to expectations but also providing empirical information that speaks to how decision makers might improve the implementation process. In the outcome stage, *theory-driven outcome evaluation* is a hybrid evaluation type that provides information not only about whether an intervention has worked or not but also about why and how the intervention worked. This information can be used for future program improvement.

Partnership Strategy

The final strategy presented in this section of the book is the *partnership strategy*, in which stakeholders invite evaluators to be partners in planning and implementing programs. The parties work closely together at every step, with evaluation information introduced regularly to support their effort to develop and implement a program. This strategy, and the bilateral empowerment approach that accompanies it, presents something of a challenge to the traditional foci of evaluation. Bilateral empowerment means that the participating evaluators are granted membership on the development team. Accordingly, they have direct input to how development and evaluation issues are handled; that is, evaluators participate in the decision-making process. Bilateral empowerment may work best with programs that have vague notions about goals, interventions, and implementation. This strategy and this approach have gained momentum in the literature concerning community coalition evaluation (e.g., Goodman, Wandersman, Chinman, Imm, & Morrissey, 1996).

APPLYING THE TYPOLOGY: STEPS TO TAKE

The purpose of the comprehensive evaluation typology is to associate particular evaluation strategies and approaches with particular program stages and stakeholder needs (see Table 2.1). The stakeholders of a program in its initial implementation stage, for example, need an evaluation strategy and approach that move quickly to tackle immediate implementation problems. The typology demonstrates a very clear truth: that *program evaluation is situational*. No single evaluation strategy, approach, or method can succeed with every possible evaluation need or situation. Means of evaluation that are fruitful in one case may be fruitless—or even misleading—in others. The performance assessment strategy, for instance, although plainly useful when the need is for accountability of a program in its mature implementation stage, could produce questionable results if employed with an immature program. This is because the only input stakeholders can actually use early on is timely information that helps to stabilize early implementation.

The comprehensive evaluation typology as it appears in Table 2.1 was crafted as a “map” of the art of evaluation for evaluators and stakeholders to review together. An evaluator might want to proceed through the typology with stakeholders, identifying the evaluation approaches and strategies best suited to the evaluation the stakeholders seek. Taking the following steps *in sequence* should bring the evaluator to the finish line in good shape.

1. *Identify the program stage that is of interest.* Stakeholders usually express evaluation needs in general, abstract terms. The evaluator must create precision in the discussion by facilitating a choice about exactly which program stage(s) should be the focus of investigation. When stakeholders request evaluation of a program implementation, they must decide whether they mean its initial implementation or its mature implementation, because the two are not the same. Lack of expressed *stage-specific* needs, understood by both stakeholders and evaluators, can end in the choosing of mismatched strategies and approaches, producing a useless evaluation. Stakeholders cannot be blamed for misunderstandings about evaluation needs because it is the evaluator’s responsibility to thoroughly grasp stakeholders’ intentions *before* designing an evaluation. Information obtained in the course of articulating or clarifying stakeholders’ needs will advance the effort to select the best evaluation strategy and approach for the task.

2. *Choose an evaluation strategy and approach that match stakeholders’ internal/external purposes.* Having settled the issue of program stage, the

evaluator must quiz stakeholders about the eventual audience for the evaluative information. Does it have an internal purpose, external purpose, or both? This is crucial when selecting an evaluation strategy. In general, if the information mainly will be used *internally* to find and fix implementation problems, then the development facilitation strategy is a good choice. For example, stakeholders desiring to troubleshoot their programs will find constructive evaluation to be valuable. For an audience outside the program, however, evaluators and stakeholders might use an assessment strategy, because assessment strategies provide much information that satisfies accountability requirements. A performance assessment strategy used at the outcome stage, for example, can be used to rigorously assess the effects of a program. But should the stakeholders need evaluative information that serves program improvement needs as well as accountability needs, then the enlightenment strategy is the best choice.

With the strategy question answered, it is time to choose an appropriate, stage-specific evaluation approach (or approaches). Each strategy included in the typology is linked to one or more evaluation approaches, and each of those is, in turn, affiliated with a number of research methods. All of these options demonstrate strengths and weaknesses in terms of the basic qualities of evaluations: timeliness, rigor, thoroughness, and cost. Stakeholders must be willing to make trade-offs among these qualities, with an adequate understanding of the pros and cons of each option.

First, an acceptable compromise needs to be reached concerning the timeliness, rigor, and cost of evaluation. There is a tendency (or at least a desire) among evaluators to take whatever evaluation approach is the most rigorous. Rigorously designed evaluations with stringent methodologies are likely to be accepted by the scientific community and perhaps published in prestigious journals. However, rigorously designed evaluations with stringent methodologies are usually expensive, and stakeholders may not have the necessary funds. Similarly, rigorous designs cannot be completed quickly, and stakeholders may be working within a window that accommodates client or community needs rather than scholarly ones. To make a generalization, the evaluation approaches and research methods within the assessment and enlightenment strategies of the typology demand more scientific rigor and so take more time to finish. On the other hand, the evaluation approaches under the typology's development strategies, although they manifest a brevity that loves deadlines, also embrace "flexible" methods like the focus group, which can be construed as departing from the rigor of the established scientific standard. This book certainly endorses the use of rigorous designs and methods where and when feasible. It equally reiterates that program evaluation is an applied science.

Serving stakeholders' needs as responsively as possible must remain a paramount concern as the evaluation approach and research method are selected. Rigor is *a* major factor, not *the* major factor, for the evaluator's consideration. So, if stakeholders offer sufficient money to support rigorous designs and methodologies, evaluators should exploit this. When money or time is necessarily limited, however, evaluators should not feel compelled to advocate an evaluation approach and research method that would be a financial burden or come to its conclusions belatedly.

One trade-off may be between time and thoroughness. The methods of outcome evaluation (such as efficacy evaluation or effectiveness evaluation) are rigorous and lengthy, whereas those of outcome monitoring are less demanding. Stakeholders whose priority is highly credible and precise information about a program's effects want outcome evaluation. Stakeholders on a tight budget of cash, time, or both want something else. If they want simply some rapid feedback about clients' progress, it would be inappropriate for an evaluator to advocate an expensive outcome evaluation when less costly outcome monitoring could also provide that feedback. The evaluator's role is to inform stakeholders that such an option exists and that it represents a trade-off, but one that will conserve their time and money.

Another trade-off is between cost and thoroughness. Evaluative information can be costly; the deeper an evaluation delves, the costlier it becomes. Programs are almost always constrained by cost, and the evaluation of results is of necessity a trade-off between evaluative product and price. Stakeholders with a program in the planning stage need to realize that they can save money by seeking only an evaluator's comments on a program plan *if* they can forgo the deeper data that costlier formative research or needs assessment would provide. Of course, if evaluator comments are unlikely to shed any new light on the program plan, the stakeholders might be better off waiting until they can afford the more expensive option.

3. *Communicate to stakeholders the facts about the chosen evaluation strategy/approach and research method.* When the evaluator has determined which evaluation strategy, evaluation approach, and research method fit the assignment best, he or she must explain them carefully to the stakeholders. Stakeholders should be especially well instructed about the kind of information that will be the final product. Communication helps prevent misunderstanding between stakeholders and evaluator. It gives stakeholders an opportunity to voice any doubts about the proposed evaluation's capacity to meet their needs. (Any evidence of such doubt should cause the evaluator to reexamine the options.) Finally, free-flowing communication with stakeholders also gives

evaluators a forum for detailing the kind of support expected from stakeholders throughout the evaluation process.

EVALUATION RANGING ACROSS SEVERAL PROGRAM STAGES

Program evaluators are frequently engaged to conduct multiple-entry evaluations, that is, evaluations across program stages. Before beginning, conflicts of interest that could be incurred by such multitasking must be addressed. Generally speaking, when the various tasks all fall within the domain of the development strategies or the domain of the assessment/enlightenment strategies, conflict of interest is negligible. Evaluators can, for example, carry out evaluation activities that assist in the development of a program plan and also, later on, provide the data to facilitate program implementation. Because each evaluation is confined to one phase and thus is of a consistent nature, the evaluations complement each other instead of competing with each other. Similarly, no conflict results when an evaluator performs assessment evaluation during the implementation stage and goes on to assess the program's effectiveness in the outcome stage. The natures of the two evaluations are compatible.

Attention to conflict of interest is warranted when evaluators doing development-oriented work with programs in their early stages subsequently become responsible for assessing program performance/merit in later stages. Whether an actual conflict exists depends on the strategies and approaches involved and on whether evaluators had a direct role in the decisions made about program planning and implementation. Conflicts of interest are quite likely to occur when evaluators conduct bilateral empowerment evaluation, becoming active members of design/development teams (as in the development partnership strategy described above), then later assume responsibility for assessing program merit. A team member–evaluator is seen as having a vested interest in the program. If he or she were to declare the program successful, the credibility of the outcome could well be suspect. Following completion of empowerment-based evaluation projects, it is much better to secure new evaluation professionals to carry out any assessment or enlightenment type of evaluation.

Evaluators whose involvement in the development facilitation strategy is limited to facilitating the work of stakeholders are not prohibited from conducting assessment or enlightenment types of evaluation of the program during later stages. An evaluator's "vested interest" is not at issue in cases in which evaluators conducted needs assessment, formative research, or formative evaluation (in the program-planning stage) for the benefit of *stakeholders* designing

or developing their program. This facilitation experience is not grounds to exclude these evaluators from evaluating the program's implementation and effectiveness later on. In the same way, an evaluator who has worked to facilitate stakeholders' development of a logic model or program theory is not barred from later conducting assessments of the program. However, as a precautionary measure to protect the perceived credibility of an evaluation, evaluators in these situations need to do three things. They must first offer up for discussion and scrutiny the fact and the nature of their earlier involvement in development activity. Second, they must make it clear to stakeholders that the requirements for evaluating programs in later stages differ from requirements for development-oriented evaluations. As a final condition, they must document explicitly how they arrived at the major conclusions of their evaluations.

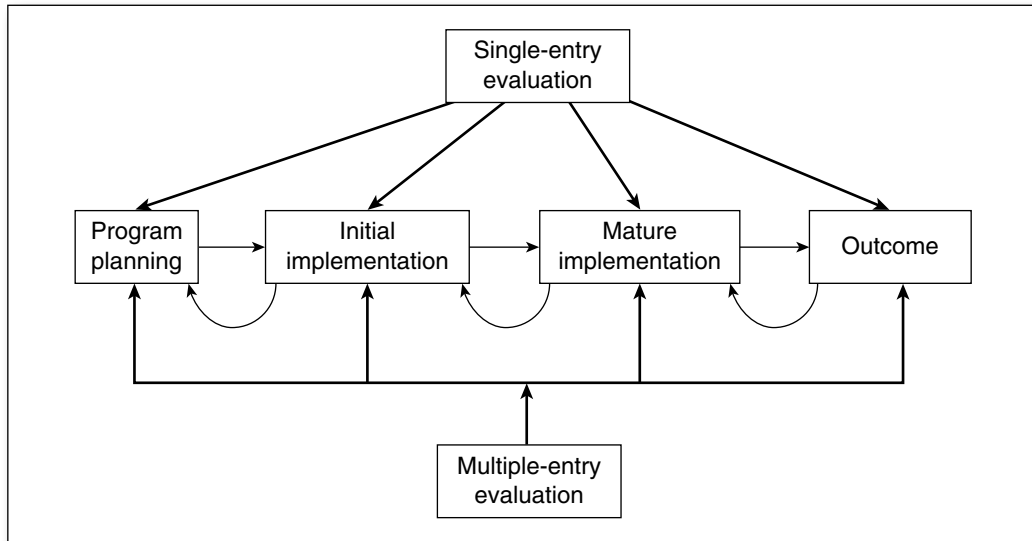
DYNAMICS OF EVALUATION ENTRIES INTO PROGRAM STAGES

The application of evaluation along program stages is dynamic in nature. Evaluators might be asked to conduct an evaluation focusing on either any one stage or a combination of stages. Figure 2.1 highlights the dynamics of such evaluation application.

The center section of Figure 2.1 indicates that program stages ideally move from planning to initial implementation to mature implementation and, eventually, to outcomes. It is possible, however, to move in a nonlinear fashion. Figure 2.1 demonstrates both single-entry and multiple-entry evaluation. The definitions of these two types of evaluation, as well as how to apply this book to conducting these two types of evaluations, are presented below.

1. Single-Entry Evaluation

In single-entry evaluation, evaluators focus their evaluation on a single program stage. This book is organized in a way that accommodates a single-entry evaluation. As long as readers have a basic knowledge of the information found in Chapters 1 through 3, they can go directly to the chapter that applies to the stage of evaluation in which they are interested. For example, if evaluators are interested in outcome evaluation, they can move from Chapter 3 directly to Chapters 9 to 12, which discuss the major issues of outcome evaluation; they do not have to refer to Chapters 4 through 8 in order to conduct an outcome evaluation, though these chapters may still prove helpful. Similarly, readers who are interested in program planning can move from

Figure 2.1 Single-Entry Evaluation Versus Multiple-Entry Evaluation

Chapter 3 to Chapters 4 and 5, which discuss evaluation approaches used in the planning stage. Readers interested in conducting an evaluation at the initial implementation stage can move from Chapter 3 to Chapter 6; readers interested in process evaluation can move from Chapter 3 to Chapter 7; readers interested in program monitoring can move from Chapter 3 to Chapter 8; and readers interested in outcome evaluation can move from Chapter 3 to Chapters 9 to 12. Chapters 13 to 15 are devoted to discussing cutting-edge issues in different program stages.

2. Multiple-Entry Evaluation

In multiple-entry evaluation, evaluators conduct an evaluation that focuses on two or more program stages. For example, at the beginning of a program, evaluators may be asked to conduct an evaluation of any two or more program stages, from planning to outcome. Similarly, in an established program, evaluators may be asked to conduct an evaluation covering both the implementation and outcome stages. This book can be used effectively to guide multiple-entry evaluations. After evaluators and stakeholders have decided which combination of program stages or evaluation approaches are to be used, evaluators can

read the chapters relevant to these stages. For example, if evaluators are asked to conduct process evaluation and outcome evaluation, they could refer to Chapter 7 and Chapters 10–12. Similarly, if they are asked to conduct evaluation in the program-planning and initial implementation stages, they could refer to Chapters 3 to 6.

The program stages in a multiple-entry evaluation could proceed in a non-linear sequence. For example, imagine that stakeholders are not happy with their existing program. They may ask evaluators to conduct an evaluation at the mature implementation stage so the decision makers can learn from their mistakes and then ask evaluators to conduct evaluations at the planning stage to facilitate the development of a new program. In this case, the evaluators could refer to Chapter 7 first and then to Chapters 3 to 5.

QUESTIONS FOR REFLECTION

1. Is program evaluation a science, an art, or both? Explain your reasoning.
2. Describe the four program stages of the comprehensive evaluation typology. Explain why evaluation needs vary across these stages.
3. What may cause a program to move back and forth among the stages, instead of moving in chronological order? Is it important for the evaluation to skip around as well, if this happens? Why or why not?
4. Why is an evaluation strategy needed? Describe four general strategies used by evaluators to guide a selection of evaluation approaches.
5. Discuss the merit assessment strategy. Give examples.
6. Discuss the development strategy. Give examples.
7. Discuss the enlightenment strategy. Give examples.
8. Discuss the partnership strategy. Give examples.
9. Why is program evaluation a situational process? Discuss why its being a situation process makes evaluation challenging.
10. Discuss single-entry evaluation and multiple-entry evaluation. How are they similar, and how do they differ? Give examples of each.
11. Discuss major challenges in conducting multiple-entry evaluation.

CHAPTER 3



Logic Models and the Action Model/Change Model Schema (Program Theory)

As discussed in Chapter 1, an intervention program is often complicated. Stakeholders need help with meaningfully describing their programs for program-planning and evaluation purposes. This chapter will introduce two tools that evaluators could use to facilitate stakeholders in developing a better description of their program. These two tools are logic models and the action model/change model schema (program theory). These two tools have their own emphases and merits. As will be illustrated, logic models are popular and relatively easy to use, and they are very useful for reducing a complicated program to a set of meaningful, manageable components. The action model/change model schema is more elaborate and takes more time to learn than do logic models. The schema is more useful when program planning or evaluation need to address contextual factors and causal mechanisms. This book encourages evaluators and stakeholders to apply either or both logic models and the action model/change model schema when facilitating stakeholders with the description of an intervention program and guiding them in evaluation design.

LOGIC MODELS

A logic model is a graphical representation of the relationship between a program's day-to-day activities and its outcomes (Julian, Jones, & Deyo, 1995; Kaplan & Garrett, 2005; Wyatt Knowlton & Phillips, 2013; McLaughlin & Jordan, 1999). Wholey (1979) rendered the logic model in two primary parts: the program components and the goals and effects

of the program. *Program components* are activities that can, either conceptually or administratively, be grouped together.

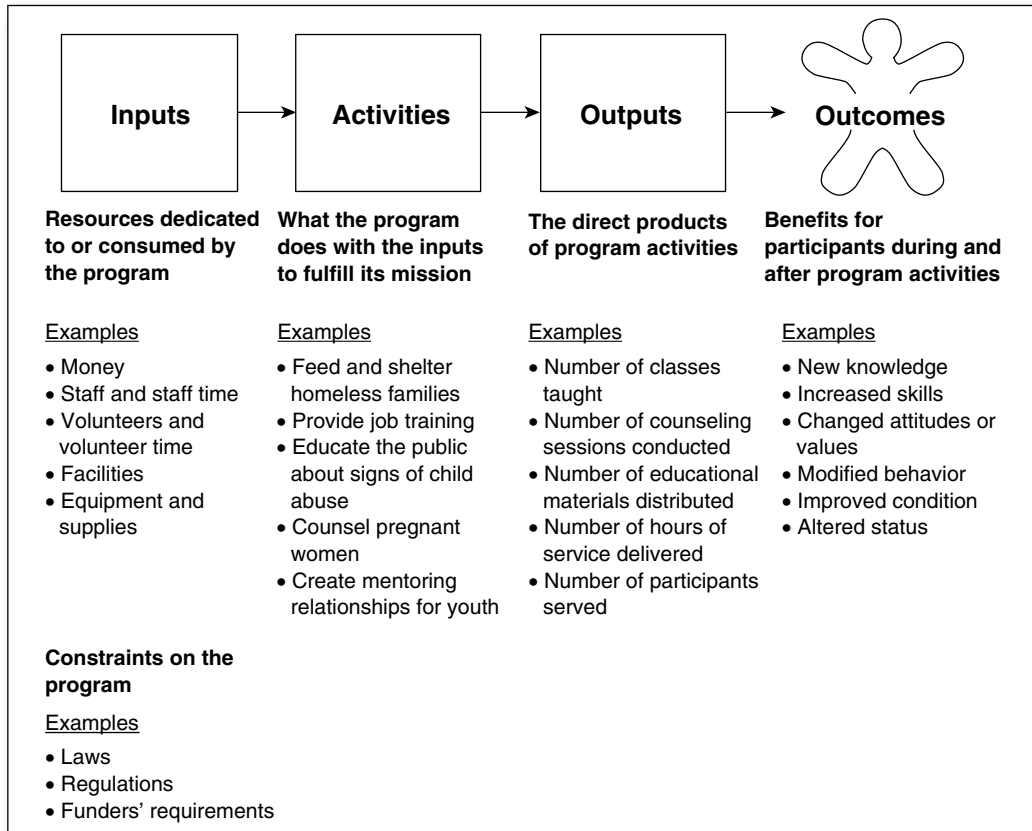
Building on Wholey's work, subsequent versions of the logic model have tended to add parts to the original. One popular twist on the model is the version developed by the United Way of America (1996). With it, evaluators of United Way programs consistently examine inputs, activities, outputs, and outcomes. In this logic model, *inputs* are defined as resources dedicated to or consumed by the program: money, supplies, staff, and even ideas. *Activities* in this model comprise services the program provides or work it performs to fulfill its mission; examples include recruiting and training staff, counseling clients, providing referral services, and educating the public. *Outputs* are defined as the direct products of program activities: number of clients served, number of classes taught, amount of goods distributed, and so on. Finally, this logic model defines *outcomes* as the benefits resulting from program activities, such as improved health, new knowledge, better skills, and higher income. These elements' relationships to each other are illustrated in Figure 3.1.

The relationships among the components in Figure 3.1 are connected by a chain of "if . . . then . . ." statements. Therefore, the relationship between the inputs component and activities component in a logic model is read as "If you have these resources as inputs, then you can use them to accomplish your planned activities." Similarly, the relationship between the activities component and outputs component is read as "If you accomplish your planned activities, then you will deliver these services or products." The relationship between the outputs component and outcomes component is read as "If you accomplished your planned outputs, then your participants will experience these beneficial outcomes."

With regard to outcomes, it is important to point out that they can occur at different levels. Some programs may focus on individual- or client-level outcomes. Outcomes at this level usually mean that participants are better off due to an intervention in areas such as knowledge, skills, finances, health, and so on. Outcomes can also happen at the group, organization, or community level. For example, a community-wide violence prevention program may target reducing violent crime rates in a community.

The basic components of a logic model discussed above can be expanded. For example, stakeholders and evaluators could expand the outcomes component into short-term outcomes and long-term outcomes. Similarly, a logic model can add a "constraints" or "external factor" component to the bottom of the figure. This component represents social, cultural, political, economic, cultural, or geographic factors that may help or hinder a program's success.

The output component of the logic model is particularly useful for monitoring purposes. For example, consider a logic model of a school-based dental care

Figure 3.1 The United Way's Logic Model

SOURCE: From United Way Worldwide Task Force on Impact. (1996). *Measuring outcome: A practical approach*. Alexandria, VA: United Way of America. Reprinted with permission.

program. The model could quantify the program's outputs, such as the number of students participating, the number of dental health brochures distributed, the number of service and education sessions conducted, and the number of schools participating. To that end, it would provide milestones for measuring a program's ongoing progress (a topic discussed extensively in Chapter 8 on program monitoring).

For a logic model to be useful, evaluators must engage the intervention program's stakeholders in its development (CDC, 1999). Stakeholder engagement allows all interested parties to reach an understanding of and agreement about program outcomes and other components. In this way, the purpose of developing

a logic model is not simply to produce a one-page diagram. Rather, the experience of participating in the model's development enhances stakeholders' buy-in to the model. This higher level of support may be key to their motivation to undertake activities outlined in the logic model.

The literature has pointed out additional merits of logic models, including the following:

- The format of logic models is frequently cited as useful for evaluators and stakeholders seeking to identify major program components and indicators (Julian et al., 1995; McLaughlin & Jordan, 1999).
- The visual presentation of a program in a logic model enhances stakeholders' understanding of program goals and resources needed for the program (Julian et al., 1995; Renger & Titcomb, 2002).

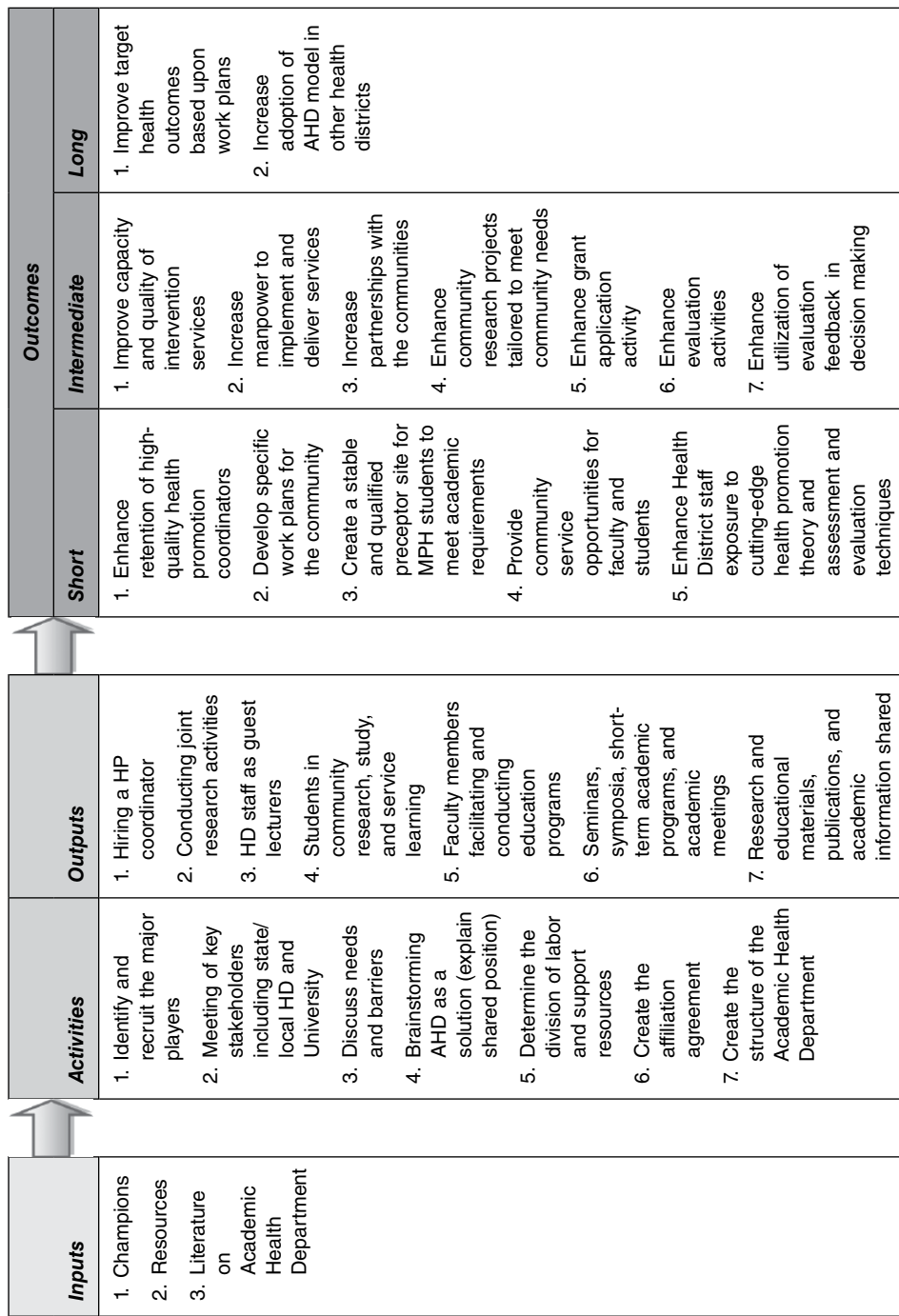
Logic models are popular in program planning. Indeed, many funding agencies require that a logic model be included with an application for funding.

Additional Examples of Applying Logic Models

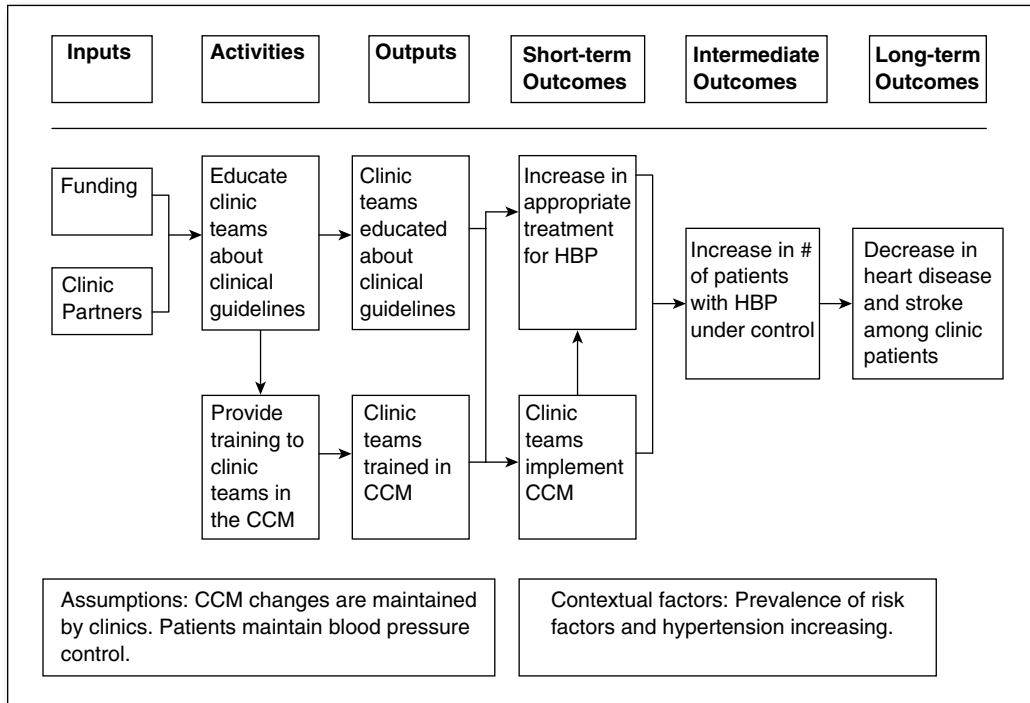
A health district in the state of Georgia and Mercer University Public Health Program entered into a formal partnership to form the Academic Health Department (AHD). The AHD would benefit both the health department, by addressing its shortage of staff, and the university, by providing students with practical learning experiences. Key partners have developed a logic model of the AHD to facilitate communication about the initiative and to guide planning, implementation, and evaluation activities (Turner, Chen, Harvey, Smith, & Redding, 2014). The major components of AHD relevant to these objectives are illustrated in Figure 3.2.

Figure 3.3 illustrates how to use a logic model to describe a CDC project to reduce the risk of heart disease and stroke. The inputs are funding and clinic partners. With these inputs, medical teams are educated about clinical guidelines and trained in the chronic care model (CCM). The outputs are teams that are educated on the clinical guidelines and trained in CCM. Short-term outcomes are the implementation of CCM and more appropriate treatment for high blood pressure (HBP). Following this, the intermediate outcome is an increase in the number of patients who have their blood pressure under control. This leads ultimately to the long-term outcome, a decrease in heart disease and stroke. This logic model assumes that clinicians will sustain their application of CCM once trained in it and that patients will sustain healthy behaviors after learning them.

Figure 3.2 Liaison-Based Academic Health Department Project: Logic Model



SOURCE: Turner, Chen, Harvey, Smith, & Redding (2014).

Figure 3.3 Logic Model to Decrease Heart Disease and Stroke

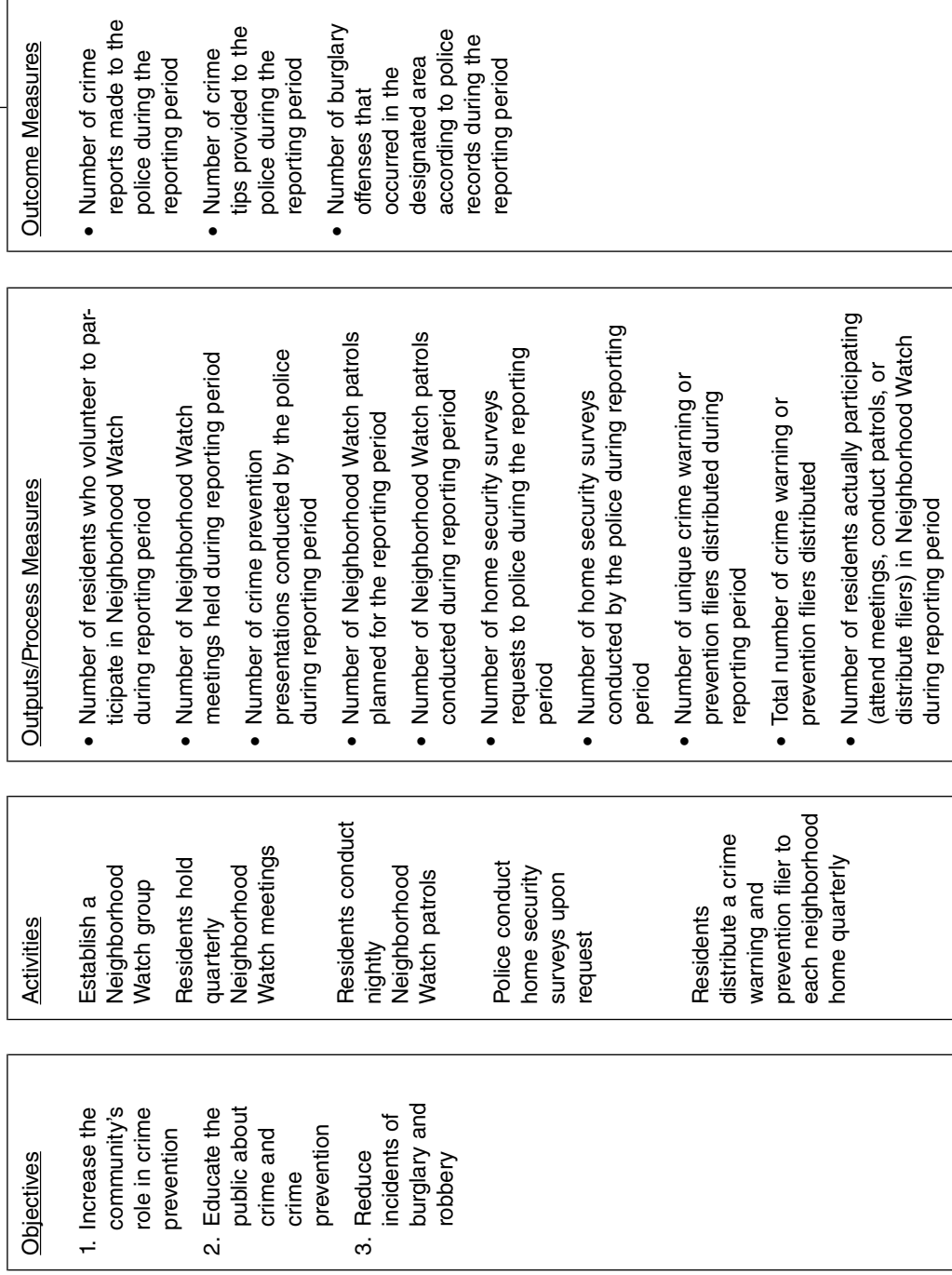
SOURCE: Centers for Disease Control and Prevention, http://www.cdc.gov/dhdsdp/programs/nhdsp_program/evaluation_guides/docs/logic_model.pdf

NOTE: CCM = chronic care model; HBP = high blood pressure.

The third example, shown in Figure 3.4, is a crime prevention program that uses a slightly different variation of the logic model. In this format, the program's objectives are listed first. The objectives state what the following activities are intended to achieve. Note that objectives are not interchangeable with inputs; inputs are not included in this logic model.

In the program modeled in Figure 3.4, the objectives are to increase the community's role in crime prevention, educate the public about crime and crime prevention, and reduce the incidence of burglary and robbery. The activities are to establish a Neighborhood Watch group, hold quarterly Neighborhood Watch meetings, conduct nightly patrols, have police conduct home security surveys upon request, and distribute a crime warning and

Figure 3.4 Crime Prevention Logic Model



SOURCE: U.S. Department of Justice, <https://www.bja.gov/evaluation/program-crime-prevention/cbcp6.htm>

prevention flier to each neighborhood home quarterly. The number of outputs and process measures is extensive. The outputs and process measures are designed to determine whether a Neighborhood Watch group has been organized that is helping residents to recognize individuals from outside their communities. If the Neighborhood Watch signs are displayed in the community and residents report illegal activity to the police, potential offenders will be deterred from committing crimes in that area. Moreover, if criminals do invade homes, they will have a more difficult time getting in and a higher chance of getting caught. Increased reporting of suspicious behavior by community members to the appropriate authorities should reduce the number of crimes that take place in the area. Overall, outcomes will be measured by tracking the number of crime reports to police, crime tips to police, and burglaries. The prior example dealing with a health care intervention distinguished among short-term, intermediate, and long-term outcomes, whereas this model does not because such a distinction is not necessary for the program it describes.

PROGRAM THEORY

Program theory is another conceptual framework that evaluators use to facilitate stakeholders in describing an intervention program or to guide an evaluation (Chen, 1990, 2005, 2012a; Chen & Turner, 2012; Coryn et al., 2011; Donaldson, 2007; Fulbright-Anderson, Kubisch, & Connell, 1998; Funnell & Rogers, 2011; Nkwake, 2013; Rossi et al., 2004; Weiss, 1998). Program theory is related to logic models but distinct from them. It emerged from the tradition of theory-driven evaluation (Chen, 1990; Chen & Rossi, 1980; Coryn et al., 2011).

The benefits of program theory for evaluation are well documented in the literature. For example, Bickman (1987) discussed the usefulness of program theory for improving the generalizability of evaluation results, contributing to social science theory, uncovering unintended effects, and achieving consensus in evaluation planning. Weiss (1998) noted that an advantage of program theory is that it provides early indications of program effectiveness. She also found program theory helpful for explaining the occurrence of program effects, thus enhancing the relevance of evaluation. In addition, Chapters 12 and 14 of the book will show that program theory can further advance evaluation knowledge and methodology by assessing and comparing the relative strengths and limitations of formal theory-based interventions versus stakeholder theory-based interventions (Chen & Turner, 2012).

One popular definition of program theory arises from causal relations. Bickman (1987), for example, defined program theory as “a plausible and sensible model of how a program is supposed to work” (p. 5). Weiss (1995) used the term “theory of change” as a way to describe the set of assumptions that explains both the mini-steps that lead to the long-term goal and the connections between program activities and outcomes that occur at each step of the way. A broader definition of program theory subsuming the existing definitions was given by Chen (1990), who described program theory as “a specification of what must be done to achieve the desirable goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated” (p. 43).

In elaborating his 1990 definition of program theory, Chen (2005) argued that the design and implementation of an intervention program are usually based on a set of explicit or implicit assumptions by stakeholders about what action is required to solve a social problem and why the problem will respond to this action. An analysis of the explicit and implicit assumptions underlying a program is essential for evaluation. Chen’s definition of program theory suggests its simultaneously prescriptive and descriptive nature, a status requiring program theory to be action oriented. Thus, program theory goes beyond typical scientific theories—those from the social and behavioral sciences, for instance—that focus solely on providing causal explanations of phenomena. Program theory can be viewed, then, as a configuration of the *prescriptive and descriptive* assumptions held by stakeholders and thus underlying the programs stakeholders create.

Program theory has different versions. This chapter will introduce a comprehensive version of program theory, called the action model/change model schema, that operationalize Chen’s (1990, 2005) definitions of program theory for practical application. Here the author seeks to explain, for evaluation practitioners, the action model/change model schema in user-friendly terms. Understanding this schema should allow practitioners to use it effectively in evaluation. Knowledge of the schema will also elucidate the how-to of applying the various approaches and methods for assessing program planning, implementation, and effectiveness discussed throughout the rest of the book.

THE ACTION MODEL/CHANGE MODEL SCHEMA

The action model/change model schema is defined as a systematic configuration of stakeholders’ prescriptive and descriptive assumptions underlying programs, whether they are explicit or implicit. Descriptive assumptions,

articulated in a change model, deal with what causal processes are expected to happen to attain program goals. Prescriptive assumptions, articulated in an action model, deal with what actions must be taken to produce desirable changes.

Descriptive Assumptions

Within the action model/change model framework, descriptive assumptions concern the causal processes underlying the social problem a program is trying to address. As an illustration, consider an intervention program for spouse abusers. According to program designers' descriptive assumptions, spouse abuse typically results, at least in part, from the abuser's lack of skill in dealing with anger or frustration and lack of knowledge of the law's stance on domestic violence. In light of these descriptive assumptions, the treatment program might be designed to employ counseling to develop anger management skills. It might also stress the legal consequences of committing domestic violence. The causal process underlying this treatment program's effectiveness, then, would be the instillation of a fear of consequences to encourage practice of the skills taught, which is then expected to reduce the abuse.

Assumptions about the causal processes through which an intervention or a treatment is supposed to work are crucial for any program, because *its effectiveness depends on their truthfulness*. If invalid assumptions dictate the strategies of a program, it is unlikely to succeed. For example, if the major motive of spouse abuse is actually belief in the patriarchal structure of families, rather than uncontrolled anger or ignorance of consequences, then an emphasis on anger management is unwarranted. The set of descriptive assumptions made about causal processes underlying intervention and its outcomes constitutes the *causative theory* (Chen, 1990) of programs. Outside the field of program evaluation, however, this phrase may not communicate well—and remember that stakeholders come from other fields. The set of descriptive assumptions can also be termed the *change model* for purposes of effective communication, and throughout this book, *change model* is substituted for *causative theory* or *descriptive theory*. The change model is emphasized in much of the theory-driven or theory-based evaluation literature (e.g., Donaldson, 2007; Weiss, 1998). As will be discussed in Chapters 4 and 5, the change model concept is very useful for providing a foundation from which stakeholders can develop a sound program plan.

Prescriptive Assumptions

Prescriptive assumptions are equally significant, according to program theory, in an intervention program. The prescriptive assumptions of program theory prescribe those components and activities that the program designers and other key stakeholders see as necessary to a program's success. Program designers' prescriptive assumptions thus direct the design of any intervention program. They determine the means of implementing and supporting the intervention so that the processes described in the change model can occur. Because prescriptive assumptions dictate which implemented components and activities will be required to activate the change model, they are collectively referred to as the *normative theory*, or *prescriptive theory*, of programs (Chen, 1990). But again, stakeholders (and evaluation practitioners, too) may appreciate the directness of an alternative term, *action model*, which is used in the remainder of this book. As will be discussed in Chapters 4 and 5, the action model concept is very useful for facilitating stakeholders in articulating the action aspect of their program plan.

Program evaluators look to the action model for the requisites of a program, as well as for the feasibility of these requisites in the field. In the action model are found the bases for answering questions such as the following: What are the crucial elements of the intervention? What kind of organization is needed to deliver the services? Who is best qualified to deliver them? How will implementers be trained? What is the target group? How will the target group be reached?

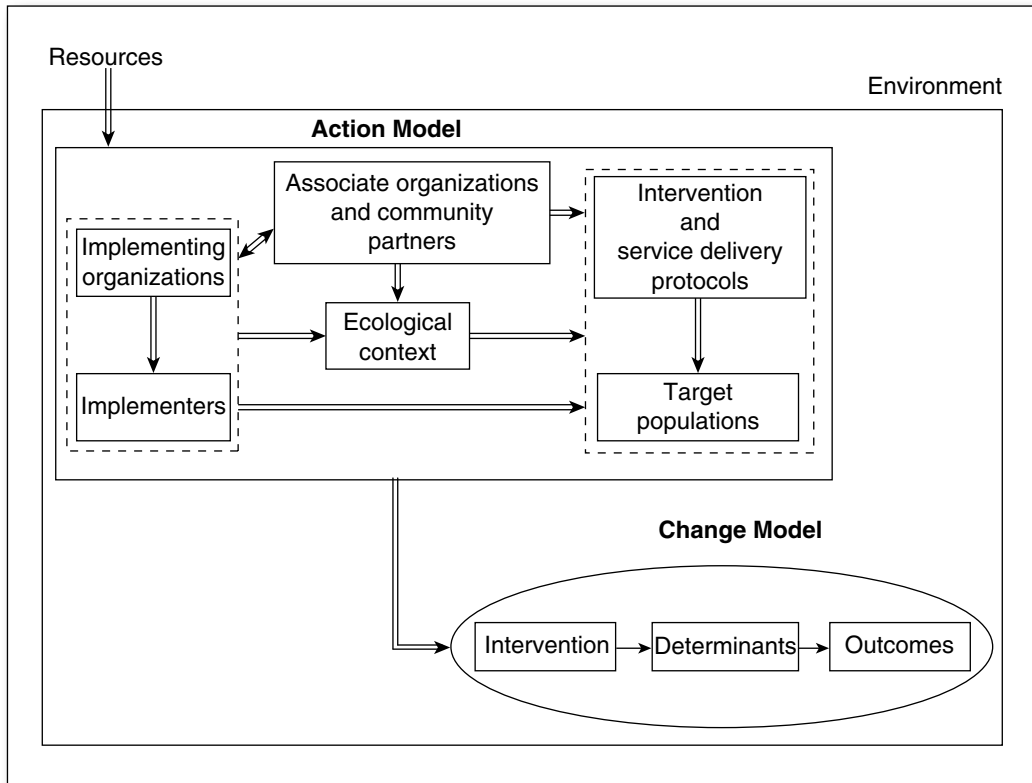
Again, as an example, take the spouse abuse treatment program. Suppose its designers decide that the target group should be abusers convicted by a court; this decision is based on an assumption that most spouse abusers end up in court and that the court will agree to use the treatment program as part of an abuser's sentence. The arrangement would certainly guarantee the program a steady source of clients. It would also necessitate establishment of an administrative linkage between the court and the program's implementing organization, based on an assumption that clear channels of communication will keep the court apprised of any client's failure to attend treatment. Suppose the program designers choose group counseling, headed by a trained and experienced professional facilitator, as the treatment for the abusers. This decision could stem from the program directors' favorable experiences with group therapy in other situations. Perhaps the designers decide that group counseling should be provided weekly for 10 weeks because they believe that 10 counseling sessions is a sufficient "dose" for most people. From these assumptions comes the need for the program to hire two professional counselors who are available for 10 consecutive weeks.

The action model deals with nuts-and-bolts issues, which are not a major topic in most modern social science theory, perhaps due to the social sciences' emphasis on developing generalizable propositions, statements, and laws. Indeed, contemporary social science theory tends to trivialize "how-to" program issues. Plus, the action model has no proposition-like format resembling that defined by and familiar to modern social scientists. However, it is interesting to note that many classic social science texts discuss both descriptive and prescriptive theories. Both Max Weber (1925/1947) and Émile Durkheim (1893/2014) intensively discussed not just explanations of organizational and societal phenomena but also steps for improving organizations and societies.

The action model translates the abstract ideas that theoretically justify a program into the systematic plan necessary to organize its day-to-day activities. Implementation of the action model puts a program in motion. And just as with the change model, if the action model is based on invalid assumptions and is thus poorly constructed or unrealistic, the program is not likely to succeed. Another example shows how important an accurate action model is to a program. The government of a developing country found that many farmers could not afford to buy fertilizer or modern equipment to increase productivity. It set up low-interest loans for the farmers. Designers of this financial program postulated a particular change model: Lack of access to capital limits farmers' ability to improve productivity, and farmers would apply for low-interest loans, if they were available, to buy machinery and fertilizer to boost their land's productivity and their earnings. The designers' programmatic model stipulated use of the government's own banks to process applications and conduct subsequent transactions. The underlying assumption was that, as part of the government system, these banks would require simply an administrative order to diligently and responsibly implement the program; in addition, operational costs would be much less than if commercial banks became involved.

A couple of years after the program had been launched, few farmers had received loans and benefited from the program. Why? Because certain assumptions of the action model were wrong. Local staff of the government bank did not see the new program as all in a day's work. To them, the program meant another burden in addition to their already heavy workload, with no increase in rewards. Consequently, the staff members' implementation of the program was not what decision makers had assumed it would be. Not only were they unenthusiastic about the program, but they also pulled up older rules and regulations to actively discourage farmers from applying for, or to disqualify them from receiving, the loans. This maintained their accustomed workload—and made the new program fail.

Figure 3.5 Conceptual Framework of the Action Model/Change Model Schema (Basic Form)



The action model/change model schema is illustrated in Figure 3.5. In the rest of this chapter, Chen's (1990) initial conceptual framework of program theory is broadened and altered, the form of the action model/change model schema, to increase its relevance within evaluation practice.

COMPONENTS OF THE CHANGE MODEL

The components of a change model are its goals and outcomes, its determinants, and the interventions or treatments it is to implement. These change model components and their interrelationships are introduced here.

Goals and Outcomes

Goals reflect the desire to fulfill unmet needs, such as may occur with poor health, inadequate education, or poverty. Program goals are established in light of certain major assumptions, such as their likelihood of being well understood and supported by staff and other stakeholders; their power to motivate commitment of resources and effort; and/or their accurate reflection of stakeholders' aims in valid, measurable outcomes. A program's existence is justified through the meeting of its goals, which are usually articulated in very general, highly laudatory language in an effort to win broad support for the program. In contrast, outcomes are the concrete, measurable aspects of these goals. For example, one goal of welfare reform is to reduce dependency on welfare. An outcome linked to this goal might be increased numbers of welfare recipients obtaining jobs, alleviating their need for government support. "Reducing dependency on welfare" is a notion with many ramifications; it is imprecise. But the outcome "obtaining jobs" gives specific meaning to the program's orientation.

Outcomes themselves may have components, and some outcomes may have both short-term and long-term manifestations. For example, in an HIV prevention program, the outcome over the short term may be increased use of condoms by a high-risk population. The outcome of the same program in the long term may be a lower number of HIV transmissions. Furthermore, a program's outcomes may include intended and unintended developments. If program stakeholders and evaluators suspect that unintended outcomes (whether desirable or undesirable) will occur, then the evaluation should include an identification of all unintended outcomes.

Determinants

To reach goals, programs require a focus, which will clarify the lines their design should follow. More specifically, each program must identify a leverage mechanism or cause of a problem, which will provide the basis of the treatment or intervention developed to meet a need. The assumption is that, once the program activates the identified leverage mechanism, or alleviation of the cause of a problem, its goals will soon be achieved. That leverage mechanism is variously called the *mediating variable*, the *intervening variable*, or the *determinant*, and in this book, the last term is used. Formal theories, developed in every discipline, provide a rich source of determinants for designing a change model. For example, in the field of health promotion, formal theories suggest a variety of determinants that program designers and key stakeholders can deploy in a program (Bartholomew, Parcel, Kok, &

Gottlieb, 2001). For example, the health belief model (e.g., Strecher & Rosenstock, 1997) outlines these determinants influencing an individual's course of action (or inaction) for a health problem: perceived susceptibility to the problem, perceived seriousness of the problem's consequences, perceived benefits of a specific action, and perceived barriers to taking action. Similarly, social learning theory (Bandura, 1977) cites self-efficacy—or the conviction that one can, in fact, carry out the behavior that elicits the outcome—as the most critical determinant of behavioral change. The PRECEDE-PROCEED model (Green & Kreuter, 1991) identifies predisposing factors, reinforcing factors, and enabling factors as important determinants for health behavior change. The determinants identified by scientific theories are intensively studied and applied in scientific research.

Of course, not many programs designed and conducted by stakeholders are intended to strictly conform to formal theories developed from the academia. Naturally, what is identified as the determinant often relates to the program designers' understanding of what causes the problem they want to alleviate and on which exact cause or causes they want a program to focus. This understanding is called *stakeholder theory*. There have been program designers, for example, who believed that urban school students' poor test performance stemmed from a lack of parental involvement, making parents the appropriate focus for programs meant to improve scores. These program designers saw in parental involvement the determinant to help students perform better; for them, it followed that, if the program activated parental involvement, student scores would improve. With a determinant identified, they could move on to figuring out how parents could be trained and motivated to help children study. Again, a program's identified determinant will provide its focus.

Social problems often have roots in multiple causes, but an intervention program usually focuses on one, or perhaps a few, determinants that program designers see as the major cause of the problem—or the most feasible to address or the one best suited to their expertise. It would be difficult for a program to deal simultaneously with all potential determinants, given typical constraints on resources and time. The unmanageability of multiple determinants aside, it remains important to specify clearly on what determinant a program has selected to focus and to justify that selection. Consider the case of juvenile delinquency in a community. High rates of such delinquency may be the result of peer pressure, failure in school, a lack of positive role models, a lack of discipline, a subculture of violence, or a dearth of economic opportunity. A program to lower rates of juvenile delinquency must state plainly, to stakeholders and the community, the cause or causes it assumes to be most relevant and the determinant or determinants upon which it will focus.

Intervention or Treatment

Intervention or treatment comprises any activity or activities of a program that aim directly at changing a determinant. Intervention/treatment is, in other words, the agent of change within the program. The vital assumption made in the intervention/treatment domain is that by implementing certain activities, the program changes the determinant and ultimately reaches its goals. For example, a treatment program for juvenile delinquency chooses to focus on a community's lack of accessible positive role models for youth. The intervention or treatment provided by the program is to team each youth with a volunteer, an accomplished professional or businessperson from the area, who will serve as a role model. Volunteers are expected to spend 2 hours each week with the participant, providing guidance and encouragement related to school, home, and neighborhood. Once a month, the pair is asked to attend a community event or visit with a private or public organization. As the pair's relationship deepens, the program designers assume, the status of the volunteer and his or her personal interest in the youth will motivate the youth to identify with the volunteer and emulate his or her agenda of productive and beneficent activities. This will lower the odds of future involvement in delinquency.

In many cases, an intervention or treatment has a number of elements. For example, alcohol abuse treatment is likely to include detoxification, individual and group counseling, and family therapy. Some intervention programs, on the other hand, can attain program goals without mediating by a determinant. Food relief programs in a disaster or warring region are a good example. A food relief program is regarded as successful as long as food is distributed to and consumed by refugees, even though the cause of their hunger, such as displacement of farmers from agricultural land or disrupted supply routes and markets, is not addressed. However, the great majority of intervention programs aim at changing knowledge, beliefs, behaviors, and/or skills. These kinds of programs usually require the intervention to change some determinants in order to affect goals or outcomes.

The terms *intervention* and *treatment* have been used interchangeably in the program evaluation literature. However, for health-related programs, at least, there is a subtle difference between the two concepts. In health-related programs, *treatment* is equal to caring for and, ideally, curing people who currently have some illness. *Intervention* more often refers to an effort to alleviate an existing problem, to ward off a potential problem, or to improve some aspect of quality of life. An intervention might sometimes comprise

treatment. The evaluation principles and strategies discussed in this book can be applied to either treatment or intervention programs. For simplicity's sake, in the remainder of the book, the term *intervention* will be used, covering both meanings.

COMPONENTS OF THE ACTION MODEL

An action model is a systematic plan for arranging staff, resources, settings, and support organizations in order to reach a target population and deliver intervention services. This programmatic model specifies the major activities a program needs to carry out: ensuring that the program's environment is supportive (or at least not hostile), recruiting and enrolling appropriate target group members to receive the intervention, hiring and training program staff, structuring modes of service delivery, designing an organization to coordinate efforts, and so on. It is vital to recognize that the impact made by a program's change model results jointly from the intervention's effect *and* the particulars of the program's implementation. The success of a job-training program, for example, is determined not entirely by its curriculum but also by the quality of its teachers, the motivation and attitude of its participants, the job search strategies employed, and the vigor of the local economy. The following discussion touches on all major elements—that is, the complete form—of the action model; it provides an exhaustive list, which may be much more than the evaluator requires in actual practice. (A rule of thumb is that large-scale programs may need all six elements, whereas small-scale programs may be just as effective with only a few of them.) Nevertheless, familiarity with the complete action model enables the evaluation practitioner to discuss more than one version of program theory. Access to the complete action model also helps in determining which components are important in a given set of circumstances and in understanding how to simplify or otherwise modify the model to fit particular evaluation needs. The elements of the action model are the implementing organizations, program implementers, associate organizations/community partners, context/environment, target population, and intervention and service delivery protocols. From this list of elements, program evaluators can draw ideas about areas of potential focus within evaluations they are designing.

Intervention and Service Delivery Protocols

The change model for a program reflects general and abstract ideas about intervention that must be translated into the set of concrete, organized, implementable

activities constituting its programmatic model. Basically, there are two requirements for this translation: an intervention protocol and a service delivery protocol. The *intervention protocol* is a curriculum or prospectus stating the exact nature, content, and activities of an intervention—in other words, the details of its orienting perspective and its operating procedures. To begin to ascertain the intervention protocol of a family-counseling program, for example, answers to the following general questions are needed: What is the nature of the counseling? What is the content of the counseling? What is the schedule for the counseling? Specific answers to these might be generated by asking questions such as the following: Is the counseling based on behavior therapy? On reality therapy? On another kind of therapy? Will counselors proceed by following standardized documents? How many counseling sessions are planned, and how long will each last?

In contrast, the *service delivery protocol* refers to the particular steps to be taken in order to deliver the intervention in the field. The service delivery protocol has four concerns: client-processing procedures, or how clients move from intake to screening to assessment to service delivery; division of labor in service delivery, or who is responsible for doing what; settings, which may be formal (e.g., at a program's office) and/or informal (e.g., in a client's home); and communication channels (face-to-face, telephone, email, website, etc.). As an example, the service delivery protocol of a program addressing child abuse would provide answers to the following questions: Where will counseling take place—in a counselor's office or in clients' homes? Will each parent be counseled separately, or will they meet with the counselor together? At what point, if any, will child and parents be counseled together? In general, one place to look for the level of quality of a program is in its establishment (or lack of establishment) of an appropriate intervention protocol and service delivery protocol.

Implementing Organizations: Assess, Enhance, and Ensure Their Capabilities

A program relies on an organization or organizations to allocate resources; coordinate activities; and recruit, train, and supervise implementers and other staff. How well a program is implemented may be related to how well the organization is structured. Initially, it is important to ensure that the implementing organization has the capacity to implement the program, and strategies exist that can be helpful in determining this. For example, if a funding agency gets to choose the implementing organization from among several qualified candidates, that agency may be well equipped to determine which organization is most capable of implementing the program. In reality, however,

such a pool of capable organizations may be missing. This is especially true for community-based organizations. Usually, an implementing organization's capacity to conduct the program must be built up. *Capacity building* involves activities such as training, transferring technology, and providing—financially and otherwise—for the hiring of experts or consultants to help plan and conduct the implementation.

Program Implementers: Recruit, Train, and Maintain Both Competency and Commitment

Program implementers are the people responsible for delivering services to clients; they include counselors, case managers, outreach workers, schoolteachers, health experts, and social workers. The implementers' qualifications and competency, commitment, enthusiasm, and other attributes can directly affect the quality of service delivery and the intervention itself. Thus the effectiveness of the program in large part depends on them. Under the action model, it is important for a program to have a plan for ensuring competency and commitment among program implementers, using strategies such as training, communication, and performance monitoring/feedback.

Associate Organizations/Community Partners: Establish Collaborations

Programs often may benefit from, or even require, cooperation or collaboration between their implementing organizations and other organizations. If linkage or partnership with these useful groups is not properly established, implementation of such programs may be hindered. In the example of the spouse abuse treatment program introduced above, program implementers need to work closely with the court to develop the procedures that will ensure convicted abusers participate in treatment as part of their sentences. This program would meet with serious difficulty if it lacked a working relationship with the court or failed to win the support of judges. Under the action model, it is important to create feasible strategies for establishing and fostering relationships with associate organizations and community partners. As will be detailed in Chapter 5, this element is most important when an evaluator is asked to take a holistic approach to help program designers and other stakeholders plan and develop a program.

Ecological Context: Seek the Support of the Environment

Ecological context is the portion of the environment that directly interacts with the program. Some programs have a special need for *contextual support*, meaning the involvement of a supportive environment in the program's work. (Indeed, most programs can be facilitated to a degree by an environment that supports the intervention processes.) A program to rehabilitate at-risk juveniles, for instance, is more likely to work when it obtains the support and participation of juveniles' families and friends. Both *micro-level contextual support* and *macro-level contextual support* can be crucial to a program's success.

Micro-level contextual support comprises social, psychological, and material supports that clients need in order to allow their continued participation in intervention programs. For example, under current welfare reform laws, in order to receive benefits, mothers must attend job training or find work. But these reforms present two immediate problems: Is transportation available to get the women to the workplace? And who will care for the children while they work? A welfare-to-work program is hardly manageable without tackling these issues. Furthermore, clients may be more likely to participate seriously in programs when they receive encouragement and support from their immediate social units (typically family, peer group, and neighborhood). When program designers or implementers realize that micro-level contextual support could play an important role in an intervention, it is up to them to try to build this support into a program's structure. For example, designers of an alcohol abuse program might organize a support group for clients that includes family members and peers who encourage and support them during and/or after intervention.

In addition to micro-level contextual support, program designers should consider the macro-level context of a program, that is, its community norms, cultures, and political and economic processes. These, too, have the ability to facilitate a program's success. A residential program for the mentally ill can anticipate real difficulties if the local community has a generally hostile attitude toward its clients. But if an adequate campaign for community support of such patients is one component of the residential program's implementation, these difficulties may be alleviated. In any case in which stakeholders believe macro-level contextual support to be crucial to their program's success, generating this support should be included as an element of their program.

Ensuring the capability of implementing organizations, establishing collaboration with associate organizations, and winning contextual support requires great effort. Finding resources to support such an effort can be a challenge. There is a worthwhile payoff, however. If a program does succeed in these

activities, it is considered an *ecological*, or *multilevel, intervention program*: It is a program with goals not just for individual clients but also for the surrounding community. Ecological programs may be likelier to attain their goals than are programs concentrating simply on client issues. This element signals a need to take a holistic approach to conduct program evaluation.

Target Population: Identify, Recruit, Screen, Serve

The target population is the group of people whom the program is intended to serve. Three assumptions that often figure in evaluation are the presence of validly established eligibility criteria, the feasibility of reaching eligible people and effectively serving them, and the willingness of potential clients to become committed to or cooperative with (or at least agreeable to joining) the program. Faced with resource constraints, a program usually cannot provide services to everyone in a target population. Therefore, it needs a clear and concrete boundary for eligibility. Criteria must also be established for determining which populations the program will recruit. For example, the target population of one Head Start program is preschool children from disadvantaged families residing in a particular community. Similarly, an HIV prevention program in one community chooses to serve addicts who inject drugs rather than trying to target the entire high-risk population. A program is usually regarded as ineffective if it finds itself serving the wrong population or failing to reach enough members of the right population. A nursing care program intended to serve low-income elderly people, for example, has failed if its services benefit many comparatively well-to-do people. Similarly, a job-training program that is well funded and well run will have failed if it produces only a handful of “graduates.”

Whether or not clients are prepared to accept the intervention also can affect program outcomes. Especially for labor-intensive types of programs, client screening and assessment are vital. A labor-intensive program must be certain of its clients’ readiness for intervention, *client readiness* being the extent to which an individual’s mental and physical state permits his or her acceptance of an intervention. If clients are not mentally and physically ready for the intervention, it is unlikely to work. *Mental readiness* of a client is the degree of his or her willingness to recognize a problem or deficiency, or the degree of motivation to accept an intervention. For example, a person who insists his or her alcohol use is not a problem will probably not succeed in an alcohol-abuse counseling program. Clients also exhibit varying degrees of *physical readiness* for interventions. Health status affects delivery of some interventions. For example, counseling

clients about HIV prevention can be difficult when they suffer from severe mental health problems or have no food or shelter. Thus, information from assessment can suggest whether a client needs services in addition to the central intervention. For example, when assessment reveals the need, program staff can refer clients for housing assistance, mental health care, education, employment, or other social services. Similarly, a client still under the influence of alcohol is no more physically ready than mentally ready for intervention. Trying to deliver alcohol counseling services is futile until the client has completed a detox program; alcohol abuse intervention starts once the client is sober.

RELATIONSHIPS AMONG COMPONENTS OF THE ACTION MODEL/CHANGE MODEL SCHEMA

It is important to understand relationships among program components. In general, program components need to be organized or connected in a meaningful way in order to achieve the intervention's goals. Figure 3.5 shows how an action model is implemented so that a change model can activate the causal process. The double-banded arrows between components within the action model represent a sequential order between these two components, such that the completion of one component provides the basis for completing the next one. For example, in the figure, the arrow from "implementing organizations" to "implementers" indicates that a capable implementing organization usually must be in place so that implementers can be adequately recruited and trained. With a spouse abuse intervention program—or virtually any program—this means that there must be an organization responsible for implementing the program before counselors or clients can be recruited. In other words, the relationships among components of the action model represent a kind of "task order" relationship: Some components must be in place and complete before others can be brought on line. The only exception is the two-way arrow between implementing organizations and associate organizations/community partners. The bidirectional arrow means that very often, the associate organizations and community partners collaborate with the implementing organizations in planning program activities; thus, they begin their involvement at the same time.

On the other hand, the solid arrows within the change model in Figure 3.5 depict causal relationships. Here, changing one element creates change in the other(s). A solid arrow leading from intervention to determinants represents the model's assumption of a causal relationship between the two. In the spouse abuse program, the model assumes that group counseling has the power to give

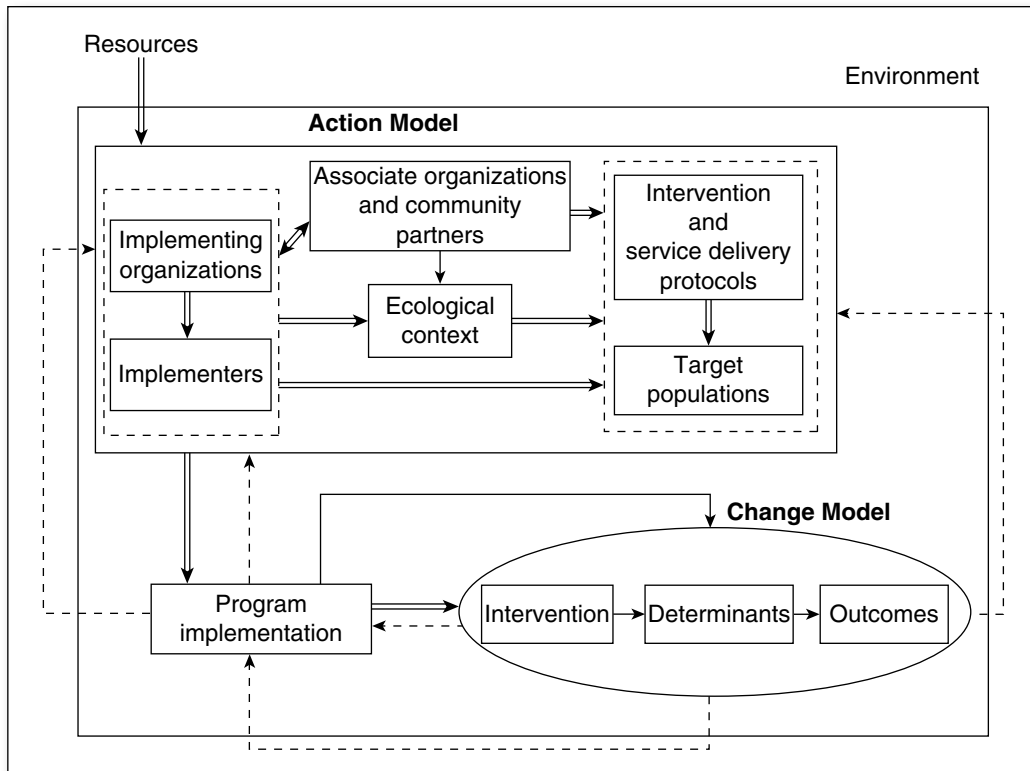
abusers anger management skills and to teach them about the criminal punishments for spouse abuse.

The schema should make clear that the action model must be implemented appropriately in order to activate the transformation process in the change model. For a program to be effective, its action model must be sound and its change model plausible; its implementation is then also likely to be effective. For example, for an HIV prevention outreach program to succeed, it needs to coordinate activities, reach the target group, and provide the group with adequate exposure to the prevention message; it must also determine which activities will strengthen the target group's knowledge of risk prevention, which should manifest itself in decreased high-risk sexual behavior. This conceptual framework of program theory should be useful to evaluators charged with designing an evaluation that produces accurate information about the dynamics leading to program success or program failure.

If evaluators and stakeholders want mainly to highlight the relationships among the components of program theory, Figure 3.5 is sufficient. However, Figure 3.5 does not address the relationships among program, environment, and feedback discussed in Chapter 1. For evaluators and stakeholders interested in elaborating these relationships, a comprehensive diagram, such as Figure 3.6, is necessary.

In Figure 3.6, the large square around the program represents its boundary. Everything within the large square is part of the program. All that is outside the square is "environment," which provides the program with necessary resources and support (in other words, its inputs) or, perhaps, works against implementation of the program. Figure 3.6 shows that, generally, a program starts with the acquisition of resources from the environment and the development of an action model. Fueled by the acquired resources, the action model can be implemented in order to activate the change model by way of the program implementation. It is the operation of the change model that leads to the attainment of program goals. A solid arrow joining an action model to a change model indicates that, strictly speaking, whatever effect the program has on the outcomes is not due solely to the implementation of an intervention but rather to a joint effect of the intervention's implementation and the implementation of other factors in the action model. Evaluation feedback is represented by dotted arrows. The evaluation feedback comprises information about how the action model was implemented in the field, such as whether the program reached the intended target population.

Similarly, the dotted arrow from the implementation to the action model indicates that evaluation feedback from the implementation can be used to improve the *planning* or the development of the action model. The dotted arrow from the change model to the implementation and action model indicates that

Figure 3.6 The Action Model/Change Model Schema (Comprehensive Form)

information from the causal process of the change model can be used to improve or modify the implementation process or the planning of the action model.

Some of the dotted lines in Figure 3.6 lie entirely inside the program boundary, while others extend outside it. These comprise two sets of evaluation feedback loops: internal and external. Internal and external feedback accommodate distinct audiences and purposes. Therefore, the evaluation approaches and strategies used with respect to the various evaluation feedback loops can be quite dissimilar. The evaluation feedback loops contained within the program boundary, the *internal feedback loops*, provide feedback for an internal audience of program implementers, administrators, and others who deal with programmatic concerns and service delivery matters on a daily basis. This audience wants from the program evaluator timely information on whether a program is operating smoothly in the expected manner. If there are difficulties, the internal audience wants to

understand, if possible, the sources of the problems as well as the likely remedies. This aspect of evaluation is called *internal-use evaluation* or *construction-oriented evaluation*. Strategies and techniques used in internal-use evaluation must be flexible and creative, and they must be accomplishable quickly. If a program is not on the right track, its course must be corrected before too much time and energy are wasted.

The other set of feedback loops in Figure 3.6 passes to the environment and then back again to the program. This set of *external feedback loops* incorporates both scrutiny by the environment and improvements from the program itself. Conducting external feedback evaluation requires more resources and more time than conducting internal feedback evaluation. The audience for external feedback is funding agencies; decision makers; interested groups; the public at large; and the stakeholders who work inside the program, such as program directors and implementers. The set of external feedback loops represents a mechanism that delivers to the environment information about the merits of a program, what changes the program may need, and the appropriate general future direction the program should take. There are two types of evaluation relating to the external feedback loop. One is intended to serve accountability needs and is called *conclusion-oriented evaluation*. The other is designed to serve both accountability and program improvement needs and is called *enlightenment-oriented evaluation*.

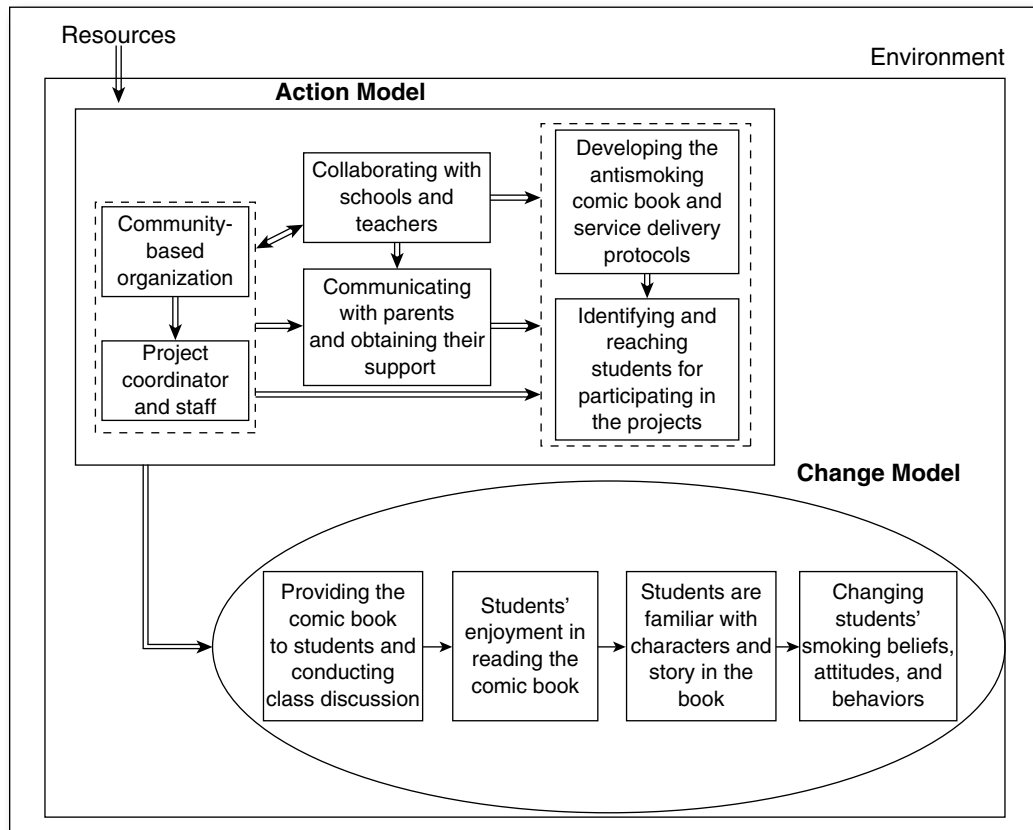
These different types of evaluation will be discussed in detail in the remainder of the book. Furthermore, the sequences of components in the action model as shown in Figure 3.6 are for general programs. The sequences can be modified according to the nature of a program. Chapter 13 will illustrate some of the variations.

APPLYING THE ACTION MODEL/CHANGE MODEL SCHEMA: AN EXAMPLE

A good example of the application of the action model/change model schema for program evaluation is found in an evaluation of an antismoking program (Chen, Quane, & Garland, 1988). Program designers devised a comic book with an antismoking story as an intervention to change students' knowledge, attitudes, and behavior concerning smoking. Program designers expressed a desire for an outcome evaluation that would provide information needed to make improvements to the program. The action model/change model schema supporting the program was stakeholder theory, stemming from the program

designers' own ideas and experiences. Evaluators conducted intensive interviews to clarify the stakeholder theory.¹ The stakeholder theory is illustrated in Figure 3.7.

Figure 3.7 The Action Model/Change Model Schema of an Antismoking Campaign



SOURCE: Adapted from Chen, Quane, & Garland (1988).

¹ How to clarify stakeholders' program theory will be discussed intensively in Chapters 4 and 5.

Change Model

The program designers' main idea for the program came from their observation that teenagers are fond of reading and collecting comic books. Accordingly, they thought a comic book that conveyed an antismoking message would create an opportunity for students to learn the facts presented about smoking and change their attitudes and behavior concerning this habit. More specifically, the program designers' change model contained two determinants in a sequential order: the students' enjoyment of reading comics and students' familiarity with the characters (heroes and villains) and story. The designers hypothesized that these determinants would lead to stronger antismoking beliefs and behaviors.

Action Model

The program designers had in mind a story, characters, and even scripts, and they collaborated with a community-based organization to implement the project. They proposed hiring a comic book artist to draw the pictures and a project coordinator and staff to run the program. They named a target population— young people attending middle school—and sought support from principals, teachers, and parents in encouraging students to participate. They planned to distribute the comic book in health classes.

After the evaluation was conducted, results showed the program to be well implemented in terms of the proposed action model. Results for the change model were mixed, however. Although students read and kept the comic book, as expected, these determinants alone did not translate into attainment of the intervention goals. Fortunately, the evaluation showed where the program's change model had misstepped, and this information would help program stakeholders design a better program.

SOME ADVANTAGES OF USING THE ACTION MODEL/CHANGE MODEL SCHEMA

Facilitation of Holistic Assessment

Using the action model/change model schema to develop contingency principles offers several advantages in the design and conduct of an evaluation. First, the conceptual framework facilitates a holistic approach to evaluating the merits of a program. Following the conceptual framework, an evaluation can explain how and why a program achieves a particular result by illustrating its means of implementation as well as underlying mechanisms that influence it.

Let us look at another example: A new curriculum has been introduced in a school in the hope of raising students' test scores. By proceeding from the conceptual framework, the evaluation of the new curriculum will do three important things: obtain information about achievement of goals, ask how effectively the action model was implemented, and explore the role of any underlying causal mechanisms. Keeping the conceptual framework in mind, the evaluator will be prompted to document the curriculum's implementation, how the students were recruited, and how the teachers taught the curriculum and were motivated to use it. The conceptual framework also prompts queries about underlying causal mechanisms: Are achieved goals truly attributable to innovations in the curriculum? Or have goals been reached by "teaching the test" to students or by taking a punitive approach to low scorers? Because the conceptual framework addresses issues in both the action model and the change model, it helps the evaluator achieve a balanced, comprehensive view of the worth of a program. This kind of assessment of what works and what does not work prevents "throwing the baby out with the bathwater."

Provision of Comprehensive Information Needed to Improve Programs

An evaluation that examines how a program's structure, implementation procedures, and causal mechanisms work in the field, as suggested by the schema, will provide information that can be very useful in program improvement. For example, if the government of a developing country wants to offer low-interest loans to farmers for costly machinery or fertilizers, program evaluators can work with program designers and other key stakeholders to critique the assumptions that underlie their work. For example, will farmers be well informed about the terms of the loans? Will local loan officers welcome a new loan program and do their part to solicit and approve loan applications under it? Evaluation of a program's underlying assumptions brings to light information that helps key stakeholders see why the program is likely or not likely to work well (or did or did not work well).

Delineation of a Strategy to Consider Stakeholders' Views and Interests

Evaluators such as Patton (2011) and Fetterman, Kaftarian, and Wandersman (2015) have argued forcefully that the design of an evaluation suffers without adequate input from stakeholders. Earlier in this section, it was noted that the schema requires evaluators to be familiar with stakeholders' assumptions about their program theory, whether these assumptions are science based or

based on personal beliefs and experiences. When stakeholders' interests and views are given due consideration during the program's design process, the evaluation's relevancy and usefulness burgeon.

Flexible Application of Research Methods to Serve Evaluation Needs

In taking a contingency approach toward research methods, the action model/change model schema offers a guide to the flexible application of methods, allowing particular evaluation issues to be effectively addressed. Few programs can be truly called identical. They all vary in structure, processes, maturity, environment, and stakeholder needs. Research methods should be tailored to meet evaluation needs, *not* vice versa. By taking on conceptual issues, the evaluator is liberated from the rigidity—the dogmatism, even—of method-driven evaluation and its ironclad research methods.

Aid to Selecting the Most Suitable Approaches or Methods

Intense conflicts among evaluators have existed over such major theoretical issues as the nature of evaluation and the chief end of evaluation, as well as over pragmatic matters such as the best methods available to our field. These persisting conflicts may have created confusion. By taking a contingency approach, the action model/change model schema ensures that the merits of a principle, strategy, or method are judged individually and in context, rather than absolutely. In context, each evaluation principle or method is granted its distinct value, not its value relative to that of competing principles or methods, and evaluators are freed to weigh them all. In this way confusion, not the options available to the evaluator, is minimized. The schema also helps make the number of options manageable by identifying those circumstances under which certain concepts and techniques are most appropriate. The schema, then, has at its heart the importance of situational factors for evaluation. This contingency view has the potential to narrow the gap between evaluation theory and evaluators' practice.

HELPING STAKEHOLDERS GEAR UP (OR CLEAR UP) THEIR ACTION MODEL/CHANGE MODEL SCHEMA

As the comprehensive evaluation typology discussed in Chapter 2 suggests, when evaluators set about reviewing a program using an approach associated

with development facilitation strategy or enlightenment strategy (strategies discussed further in the chapters that follow), a frequent first requirement is clarification of the stakeholders' action model/change model schema (Chen, 2003, 2005) or logic model, if evaluators choose to do so. At times, the evaluator may even need to help the stakeholders with the initial draft of an action model/change model schema or logic model. This section explores ways to clarify or help develop stakeholders' program theories. These strategies and techniques discussed below are also applicable to develop logic models.

Reviewing Existing Documents and Materials

To start the process, evaluators need to study existing documents or materials related to the program—brochures, pamphlets, grant applications, memos, and so on. This general information prepares the evaluator for subsequent interviews with stakeholders, ensuring that these will be conducted efficiently. Evaluators might also consider visiting program sites to increase their familiarity with programs that have already been implemented.

Clarifying Stakeholders' Theory

As the evaluator begins to clarify stakeholders' program theory, or as stakeholders begin to develop such a theory with assistance from the evaluator, an important issue must be resolved: What role should the evaluator play in this process? How can he or she best contribute to the work? The evaluator should remember that an action model/change model schema *belongs to the stakeholders*; the evaluator's function is that of facilitator and consultant. Evaluation skills and knowledge should be brought to bear to increase the productivity of the meetings at which various stakeholders attempt to articulate and refine their ideas about the program theory. Stakeholders are sure to have divergent backgrounds, concerns, and interests. It is easy for them to eat up time with free-form discussions that never even approach agreement. The evaluator's job as facilitator is to outline for the group the salient issues to discuss, showing stakeholders where to fill in with their own experiences, thoughts, and expertise. Next, the evaluator can synthesize the discussions and build consensus. The evaluator's concurrent job as consultant means filling in with his or her own evaluation expertise when stakeholders ask for advice. The evaluator is present to lay out options for stakeholders to consider and should avoid imposing his or her own values upon stakeholders. The evaluator should also present ideas drawn from his or her own expertise for stakeholders to discuss.

Participatory Modes for Development Facilitation

Evaluators can assist stakeholders whose action model/change model schema is under development by adopting either of two general participatory modes: the intensive interview mode or the working group mode. Choosing a mode is a prerequisite for stakeholders and evaluators preparing to work together.

The *intensive interview mode* centers on individual, intensive interviews that the evaluator holds with representatives from each key stakeholder group. The aim is to record systematically the individuals' perceptions about issues within the incipient program theory. Based upon these interviews, the evaluator formulates a first draft of the action model/change model schema, which the representatives and other stakeholders will then read. Their comments are incorporated into the final draft. In addition, evaluators can meet with these individuals for the purpose of fine-tuning and finalizing the program theory.

The *working group mode* similarly involves representatives from key stakeholder groups. However, in this mode, the representatives are not interviewed individually but instead meet together with the evaluator to develop the program theory. Group members need to include those who will be most deeply involved in formulating and designing the program, those who will be most deeply involved in implementing the program, and other key constituencies whose input will be influential as to the program's direction. The facilitator, of course, is another member.

The working group actually has relatively few participants when the planned program is a small one. With large programs, however, the working group easily becomes too large to work effectively. A group that is too large can discourage members' full participation, at the same time necessitating many more sessions to finish the work. A good rule of thumb is to limit a group to no more than 15 members. Small groups can foster a casual atmosphere for discussion, enabling the evaluator to serve as both facilitator and consultant. A large group, especially one with a highly diverse and vocal membership, makes it difficult for the evaluator to be facilitator and consultant at once. With large groups, at least two evaluators may need to participate in the meetings—one as facilitator, the other as consultant.

How should one choose a participatory mode? Each has its advantages. The intensive interview mode tends to be less challenging logistically because group meeting arrangements are needed only infrequently. In addition, the interview setting may strike some participants as being much more comfortable and secure than a typical meeting. The interview also tends to allow the evaluator

to more readily probe stakeholders' views. A potential limitation of the intensive interview mode, however, is some stakeholders' perception that they have participated in only one part of the theorizing process. This is especially problematic in large programs with many powerful stakeholders.

In contrast, the working group mode tends to demonstrate that the action model/change model schema is being developed in an open, inclusive manner, which could increase some stakeholders' buy-in. But again, work with a group often requires more time to finish than work done in interviews. Furthermore, it is possible in working groups for a few highly vocal stakeholders to dominate the discussion. This problem might be alleviated if the evaluator sets clear ground rules for discussion during the first meeting that encourage full participation by all members. An even more serious problem with the working group mode is that some stakeholders—those in the lower ranks of the implementing organization(s)—may worry about expressing their actual opinions, choosing instead to simply echo what higher-ranking officials say. In such a case, the final action model/change model could reflect only the views of those in authority. If this is a concern, the intensive interview mode is the better choice.

Theorizing Procedures for Development Facilitation

As with the participatory mode, a *theorizing procedure* must be selected in order to help stakeholders develop their action model/change model schema. So-called forward reasoning, backward reasoning, and forward/backward reasoning are the three general options for evaluators working within the development strategies.

Backward reasoning begins with the change model, then moves backward step-by-step to the action model in order to obtain the action model/change model schema. It is “backward” reasoning in that the process moves in the opposite direction as the sequences shown in Figure 3.5. More specifically, backward reasoning starts from the question of what goals the program seeks to achieve. Other questions are the following: On which determinants of these goals should the program focus? What intervention will affect these determinants in appropriate ways? When a change model has been completed, evaluators can facilitate stakeholders' development of the corresponding action model with questions such as these: Which groups need to be reached and served? What kind of program implementers and implementing organizations will suit? What types of intervention and implementation protocols seem best? Should there be collaboration with other organizations? Will the program require ecological support?

Forward reasoning, on the other hand, means formulating an action model/change model in accord with the logic flow outlined in Figure 3.5—action model first, then change model. Forward reasoning produces general program goals and grows from initial thoughts about what kind of action model is needed. Questions like these are important in forward reasoning: At which intervention and implementation protocols will the implementing organizations excel as they try to solve particular problems or reach certain goals? What group needs to be reached with the intervention, and *how* can it be reached? What setting and delivery mode make sense? Do clients face barriers to receiving services, and can the program alleviate these? How and where should contextual support be sought for the intervention, if needed? When they have completed the action model, evaluators and stakeholders can develop a change model by asking two questions, in sequence: What determinants will be changed by the intervention? What outcomes will be achieved by changing these determinants?

Forward reasoning and backward reasoning alike can be used successfully in the formulation of program theories. In certain circumstances, however, one of the two theorizing procedures is clearly the better choice. Some rules of thumb can guide the evaluator.

The first rule says that, generally speaking, when program designers and other key stakeholders are familiar with social science methodology, backward reasoning works best. It is the procedure that starts with discussion of a program's goals, a subject stakeholders enjoy discussing and that can help break the ice. Subsequent inquiries within the backward-reasoning procedure (e.g., What are the causes of the problem? Which intervention seems to offer promise? What is an appropriate design for the intervention?) are well within the stakeholders' capability to debate. On the other hand, when program designers and other key stakeholders are not familiar with social science methodology, forward reasoning should be preferred. The reason is that theorizing procedures need to start with a topic that stakeholders feel comfortable discussing. Forward reasoning starts with the specification of programming issues, about which stakeholders have many ideas to voice. Forward reasoning aptly suits efforts to clarify or develop stakeholders' views on the steps their program should take: what to do first, what to bring in next, building up to the third and fourth and fifth steps, and so on through culmination in delivery of a service or services. Whether an evaluation begins with forward or backward reasoning, if the evaluator and stakeholders come to realize that continuing in that mode will be difficult, they are always free to switch to the other procedure to resume their discussions.

It is also important to note that forward and backward reasoning are not mutually exclusive. In fact, *forward/backward reasoning* is a use of forward and backward reasoning, back and forth, to facilitate stakeholders and make explicit their action model/change model schema. Forward/backward reasoning is more time-consuming than the other two approaches, but it may bring to bear the best of both worlds. In using this technique, evaluators and stakeholders often apply backward reasoning first and then use forward reasoning to compensate for weaknesses in backward reasoning. For example, an evaluation focused on *both* action and change models might begin with the forward reasoning procedure to construct an action model, take up backward reasoning to establish a change model, and finally integrate the two to arrive at an overall program theory. This dual procedure is a good choice when program stakeholders and evaluators believe that unintended outcomes will be of import. Employing the theorizing procedures in both directions may make it more likely that a working group will be alerted to potential unintended desirable or undesirable effects. The evaluator should facilitate discussion of any unintended effects and their prevention, should they be undesired.

Preparing a Rough Draft That Facilitates Discussion

The work of developing a useful action model/change model schema is often time limited. The program theory's usefulness may dwindle with the passing of a deadline, and, more often than not, deadlines come sooner than the planning team would like. Scheduling, preparing for, and executing either interviews or meetings, and then compiling the information obtained and soliciting comments on it, is very time-consuming (especially so if every element and issue needs to be broached, examined, and ruled on—from scratch—in these meetings or interviews). To shorten the period required, it is not unusual for evaluators to scour existing information about a program and use what they learn to prepare a rough draft of a program theory for discussion by the working group. The rough draft should include the elements of a program theory stated in the existing information, the elements that may be implicit in the existing information but are not communicated straightforwardly there, and the significant elements not yet touched on that will require intensive discussion. The rough draft provides a focus for stakeholders' thoughts and suggestions. It should be distributed to members of the working group (or to individuals scheduled for interview) well in advance of the meeting date, giving them time to digest the contents. The rough draft is a tool to streamline discussion, focus comment, and foster specificity and usefulness in the work.

APPLICATIONS OF LOGIC MODELS AND THE ACTION MODEL/CHANGE MODEL SCHEMA

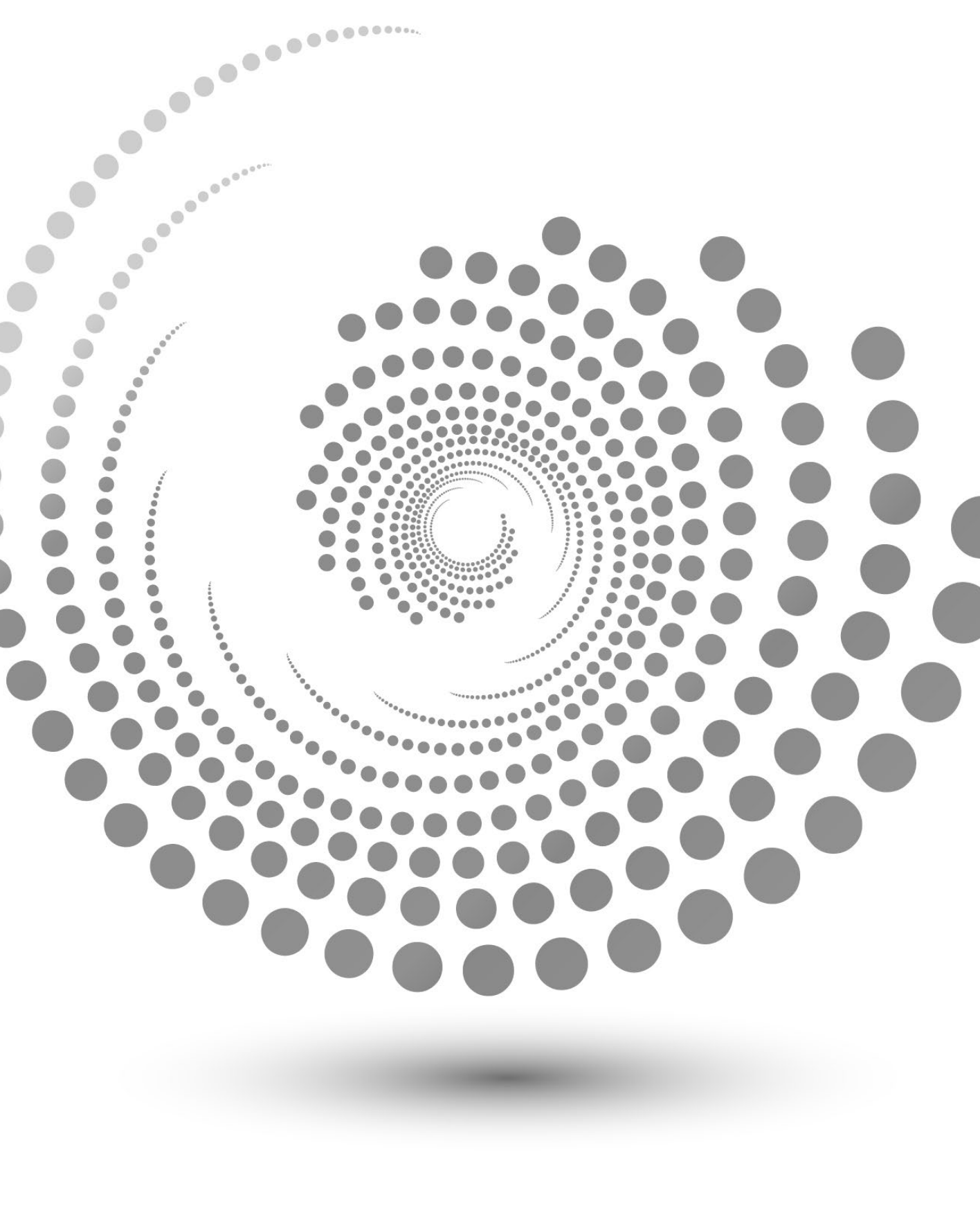
Both logic models and the action model/change model schema are useful to evaluators. Stakeholders often ask evaluators to work on logic models for their programs because of grant application requirements. A straightforward application of logic models may work for many programs. However, evaluators should be aware that such a straightforward application of logic models may not work well for programs with an emphasis on contextual factors and/or causal mechanisms. Chapter 12 will discuss the problems created by such applications and will show how they can be resolved by bringing in the action model/change model schema.

Because of its comprehensiveness, this book will emphasize how to apply the action model/change model schema so as to design and conduct fruitful evaluations that assess program planning, implementation, and/or outcomes. As the reader progresses further into this book, the three general purposes of the action model/change model schema are explained. One purpose of the schema is to underpin the comprehensive evaluation typology discussed in Chapter 2. A second purpose is to lay out for the practitioner those evaluation approaches best suited to the *program planning*, *implementation*, and *outcome* stages, suggesting some applications for these approaches. A final purpose of the schema is to use it as a platform for introducing conventional and cutting-edge evaluation approaches to evaluators and stakeholders.

QUESTIONS FOR REFLECTION

1. Use the components of logic models to describe a real-world program.
2. Using components from your example in question 1, create “If . . . then . . .” statements.
3. Why would you want to separate outcomes into short-term and/or long-term outcomes? Give examples of instances in which you would do this.
4. Describe the descriptive and prescriptive assumptions in a program theory. Compare and contrast program theory with behavioral or social science theories (formal theories).
5. Describe the components of a change model. Use a real-world program to illustrate these components.
6. Why is it important to identify the determinants of a problem? What are the possible consequences if these determinants are not defined?

7. Describe the components of an action model. Use a real-world program to illustrate these components.
8. Give examples of how a real-world organization ensures implementers' competency and commitment. What might be the consequences if the organization did not do so?
9. Can you think of any programs that went awry at least in part because the program implementers' capability was compromised? How might this have been avoided?
10. How does the intervention protocol differ from the service delivery protocol? Give examples of both.
11. Explain how the ecological context interacts with the program. What could happen if the ecological context was unsupportive of the program?
12. What conditions must hold for each theorizing procedure option to be effective? For example, why would you employ the backward-reasoning option instead of the forwarding-reasoning option? Under what circumstances would the option chosen not make a difference?
13. You are planning a program to enhance the grade point average of children from low-income families. Describe your program using both a logic model and the action model/change model schema. Compare and contrast how your logic model and action model/change model schema represent your program.
14. If an evaluator facilitates stakeholders in developing a logic model or program theory, would the evaluator's objectivity necessarily be compromised during the evaluation? Explain why or why not.

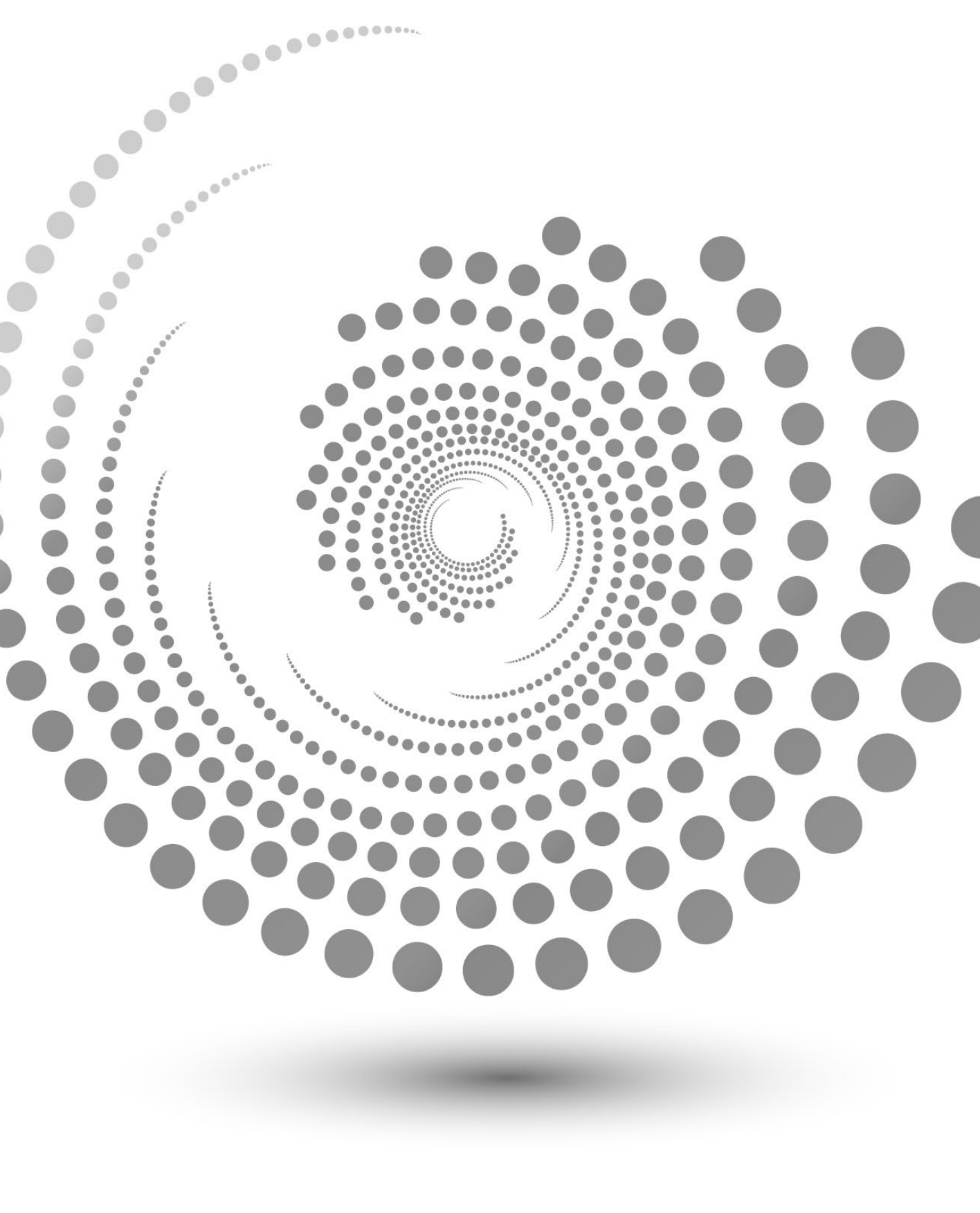


PART II

Program Evaluation to Help Stakeholders Develop a Program Plan

Program evaluation is now much expanded from traditional areas such as outcome and implementation. With a growing awareness that the implementation and outcomes of programs are affected by the quality of program planning, a further role was revealed for evaluation. The strategies and approaches of program evaluation foster the kind of understanding most likely to ensure top-quality program planning.

Generally speaking, the earlier that program evaluation techniques are incorporated into the planning of a program, the easier it becomes for the directors and implementers to improve the new program using evaluation feedback. Program staff can modify a program much more readily during the planning stage than later on. Once a program is established and on its way to becoming routine, enacting substantive changes can be difficult, even when evaluation results strongly support them. Program evaluators who are serious about putting their evaluation results to work need to learn how to apply evaluation strategies and approaches to assist stakeholders in program planning and development. The first requirement of a sound program plan is the well-developed program scope, which is the topic of Chapter 4. Chapter 5 takes up the preparation of the action plan itself.



CHAPTER 4



Helping Stakeholders Clarify a Program Plan

Program Scope

The design of an intervention program is the responsibility of program designers and other key stakeholders. Knowing how important—and complicated—the planning of a program is, these parties will frequently seek expert help with the planning process. This happens, of course, at the program planning stage (as illustrated by the comprehension evaluation typology shown in Table 2.1). This chapter was written with start-up programs uppermost in mind, but its content has implications for established programs as well. Because it is not unusual for established programs to experience changes in policy, clients, personnel, management, and/or leadership over the years, stakeholders may periodically feel that they have lost sight of their program and need to reconceptualize it, with help from the evaluator.

THE PROGRAM PLAN, PROGRAM SCOPE, AND ACTION PLAN

Program planners or designers have their own planning concepts, and they have their own terminology such as program plan, program scope, and action plan. This chapter introduces how to apply evaluation approaches and tools to help stakeholders strengthen the quality of these elements. The *program plan* is a systematic description of what the problem is; why it is important to address the problem; what needs to be done to address it; and what and who will do what, when, and where. A program plan consists of two major

parts: the program scope and the action plan. Evaluators can facilitate stakeholders in clarifying the program scope and action plan.

The *program scope* is stakeholders' statement of program focus and goals. It includes stakeholders' reasons for selecting these goals and target population. It also states the scope of the intervention to be used and explains how it is expected to lead to achievement of the goal. In addition, the program scope specifies the target population to be serviced by the intervention and explains why the population requires it. The *action plan* details how the intervention will be implemented, given the goals for the target population, and the boundaries set by the program scope. In other words, the program scope issues a call for action, and the action plan is a blueprint of that action.

The interwoven nature of the program scope and program plan means that the two are usually developed together. Nevertheless, they do differ conceptually. The program scope focuses more on the change model, whereas the action plan is more concerned with the action model. There are, in fact, advantages to developing the program scope fully before attempting to arrive at the action plan. Note that the action model/change model schema discussed in Chapter 3 is useful for developing both the program scope and the action plan. The relationship between the schema on the one hand and the program scope and action plan on the other are as follows:

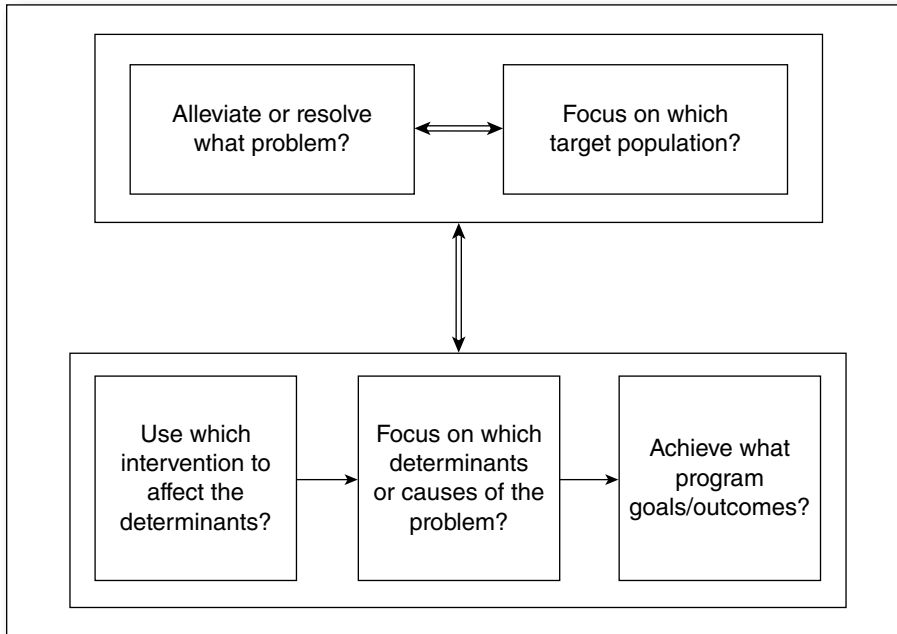
- *Program scope*. This consists of the change model plus two boundary questions (What problem is being solved? For whom is the problem being solved?).
- *Action plan*. The ideas and content of the action plan are addressed by the action model.

Program designers and other stakeholders benefit from the facilitation services provided by evaluators because ultimately the stakeholders have a coherent program plan (program scope and action plan) they can use to guide service delivery and an action model/change model that will guide evaluation activities. In this chapter, the program scope is central, whereas the action plan is featured in the next chapter.

CONCEPTUAL FRAMEWORK OF THE PROGRAM SCOPE

The conceptual framework of the program scope can be represented by a set of questions related to a program's boundary change model. Figure 4.1 illustrates a conceptual framework that typically underlies the articulation of a program's scope. In the figure, the component concepts are grouped in two boxes. The top

Figure 4.1 Conceptual Framework of the Program Scope



box includes two components—the identified problem and the population targeted for services—which define the boundaries of the program. The bottom box contains each component of the change model. Two-way arrows link the program’s boundaries and the change model to demonstrate the importance of the “fit” between the change model and the problem and target group the program seeks to address. In other words, it is vital that interventions, determinants, and goals (the three change model components) are appropriate to the target population and the problem it faces. Figure 4.1 also models the kind of diagram that is almost always helpful when the evaluator presents a finished program scope to its audience, though at times one may elect to present the scope simply as a written statement.

In summary, a sound program scope needs to cover the following set of questions:

- What problems will the program alleviate or solve?
- What target population or populations will it focus on?
- What program goals/outcomes will be achieved?
- What determinants or causes of problems will be the focus?
- What interventions will be used to affect the determinants?

Why Develop a Program Scope?

Programs rely on their program scope as a foundation for *planning*, for efficient *communication*, and as a basis of outcome *evaluation*. Program plans must be developed systematically. One major weakness in much contemporary planning of social betterment and health promotion programs is that program designers tend to create a plan in a rush. Giving the program scope but a passing glance, they proceed, even though they lack a clear vision. In consequence, their program plans tend to be unfocused and disorganized, and they are unlikely to be effective.

A well-articulated program scope can give program designers a firm foundation for their overall program plan and efforts. It is an outline they follow, a constant reminder of what makes program activities in the action model meaningful and prone to achieve the intended goals. Furthermore, the program scope can give insight into the way individual efforts support the overall program mission. A good program scope also fosters efficient communication among the program director, the program staff, and audiences within and outside of the program. It is common today for program designers to use lengthy narrative descriptions to introduce people to their programs. Unfortunately, these descriptions may take too much time to read, resulting in their intended audience simply not reading them. Furthermore, they may get bogged down in details while straying from the big-picture ideas the audience really wants and needs to know, like what the program seeks to do. In contrast, the program designer with a well-articulated program scope has at hand a concise, comprehensible summary of the program, one useful for introducing the program and facilitating communication about it. Yet another function of a sound program scope is that it establishes a basis for later evaluation of program outcomes (see Chapters 10 to 12 for an intensive discussion of this function). Containing as it does statements of the interventions, determinants, and outcomes agreed upon by stakeholders, the program scope is an excellent foundation for evaluations of various kinds.

STRATEGIES FOR ARTICULATING THE PROGRAM SCOPE

Evaluators who are invited to help stakeholders with developing a program's scope need to know that a range of evaluation strategies and approaches is available to them. Some options are *most* fruitful when used in certain limited circumstances, but it is important to realize that the available strategies and approaches are not mutually exclusive. Evaluators are free to apply a combination of them to

serve stakeholders' needs. Three strategies from Table 2.1 are explored in depth below: the background information provision strategy, development facilitation strategy, and troubleshooting strategy. The first is principally for designers and key stakeholders who are uncertain of the most pressing needs of a community or the kinds of intervention likely to be accepted by it. Using approaches such as needs assessment and formative research, evaluators systematically discover the missing information. The second strategy, development facilitation, typically employs the working-group or intensive interview approach to facilitate stakeholders' work on developing the program's scope. The last strategy, troubleshooting, may be helpful when stakeholders or evaluators believe the feasibility of an existing program scope needs to be field-tested. The plausibility-testing approach can generate preliminary information about the program scope's assumptions.

BACKGROUND INFORMATION PROVISION STRATEGY AND APPROACHES

With the background information provision strategy, evaluators can gather pertinent empirical information about community needs, target group characteristics, and clients' and implementers' perspectives on interventions. This information helps program designers devise a program's scope. The well-known evaluation approaches of *needs assessment* and *formative research* are suited to the collection of background information for stakeholders. These terms have come to be used interchangeably, but despite their similarities, each approach has a distinct focus. Because needs assessment is especially suited to identifying the unmet needs in a community, stakeholders puzzling out the goals of a program will be well served by it. Formative research is better at empirically describing clients' cultural backgrounds and capabilities; it is also useful for pinning down clients' (and implementers') opinions about proposed programs. This distinction is an important one.

Needs Assessment

Needs assessment involves the use of research procedures to identify, measure, and prioritize the needs of a community; it has been discussed intensively in the literature (e.g., Rossi et al., 2004; Witkin & Altschuld, 1995). The needs assessment strategy can facilitate selection of goals and target groups. Unmet needs in a community can be identified from existing data (e.g., census findings, vital statistics, agency records) and/or by studying a population via surveys,

focus groups, or interviews. For example, one youth agency brought in program evaluators to assess needs among African-American youths in a community so that new programs could be developed to serve this target group (Chen & Mark, 1996). The evaluators surveyed the youths and their parents and identified a set of needs, assigning each a priority. Help with schoolwork was first, drug abuse prevention second, parenthood education third, recreation fourth, and health promotion fifth. Given these priorities, stakeholders planned a tutoring program for the youths.

Formative Research

Background information can also be produced for stakeholders by using the formative research approach, which looks beyond a community's unmet needs. Formative research is a systematic method of gathering empirical information about potential clients and implementers as well as their views of proposed programs. The formative research approach investigates these individuals' characteristics and cultures, and it samples their opinions about proposed goals and interventions drawn from program scopes and plans. Popular research methods associated with the formative research approach are the focus group, the interview, and the survey. Focus groups tend to conserve time and resources, as compared to the other research methods, and they often generate more innovative ideas. Interviews tend to produce data that are more detailed and more reliable than data from other research procedures. Finally, although surveys have the advantage of including comparatively large numbers of individuals within a representative sample, the information they can provide is relatively limited.

A good illustration of the use of formative research can be found in a program in Hawaii developed to teach primary school students about preventing skin cancer (Glantz, Carbone, & Song, 1999). Planners sought program evaluation to help them understand the background of their target group as they proceeded to refine their program scope. They wanted to know more about the students' ideas and activities concerning sun protection and skin cancer; they saw the usefulness of comprehending the target group's likely response to various educational materials and sun safety policies. Group discussions, focus groups, and interviews were conducted with the children, their parents, and staff members from local recreational organizations. The data obtained indicated that sun protection practices were inconsistent, even though general awareness of skin cancer prevention was widespread. The research showed that the children did not understand what skin cancer was and they did not perceive it as a threat. This information justified the program's goals. Furthermore, data

from formative research helped these program planners choose their determinants and intervention strategies. It suggested that children were reluctant to don long pants, long sleeves, and wide-brimmed hats to protect themselves from the sun. Messages promoting this degree of covering up would appear extreme to the students—too extreme to accept. A better message, according to the formative research, would urge gradual change, beginning with wearing short sleeves, longer shorts, and caps. Research findings for the parents and recreational staff showed their concern, also, that changes be acceptable within the culture of their tropical home, where, for generations, the preferred mode of dress has been light. These adults were, however, supportive of education and policy aimed at improving their own children's sun protection habits. As to the type of materials and strategies that would reach both adult and juvenile audiences, the formative research asserted the importance of creative, engaging sun protection messages, with some to be delivered in recreational facilities and others at home. Thus, not only did the formative research help with development of the program's scope, but it also provided input for program development generally. By the way, this example points up the difference between formative research and needs assessment. Needs assessment showed planners that better sun protection was actually needed by the target population, whereas the formative research homed in on which program features were most likely to fulfill the need.

THE CONCEPTUALIZATION FACILITATION APPROACH: A PART OF THE DEVELOPMENT FACILITATION STRATEGY

Program designers and key stakeholders may have plenty of ideas for their program scope. However, they often do not know how to clarify their thoughts and connect them systematically and coherently. The evaluator who is asked to help them can turn to the *conceptualization facilitation approach*. In this approach, the evaluator facilitates stakeholders' work to conceptualize their program scope or plan. There are certain guidelines for effective application of the conceptualization facilitation approach.

Working Group or Intensive Interview Format?

To use a conceptualization facilitation approach, stakeholders and evaluators first need to select either the *working group* or *intensive interview* format (introduced in Chapter 3). The intensive interview format is typically reserved

for small-scale programs in which a few key people are charged with developing a program. Evaluators interview these key individuals with one of two aims: to clarify their program scope or to review the conceptual framework of the change model and draft a program scope. For large-scale programs, the working group format is particularly well suited because consensus among numbers of stakeholders is always a driving issue. As a strong consensus builder, the working group format can create support for the program's scope.

Theorizing Methods

Once a format is selected, the evaluator and stakeholders must determine whether forward reasoning or backward reasoning is the better theorizing method for their purpose. Stakeholders' backgrounds and preferences play a role here. Some program designers and other stakeholders may have extensive experience with and expertise in various interventions. If their minds are set on a favorite intervention, forward reasoning is the preferred method of developing a program scope. Stakeholders adept at social marketing to promote health, for example, may decide to seek funds for such an intervention in their community. Using forward reasoning, the evaluator could facilitate these stakeholders' in identifying which determinants the intervention will affect and help them estimate the eventual outcome.

With different stakeholders—say, a group that remains open to a process of selection—backward reasoning may effect a successful search for the right intervention. Backward reasoning demonstrates great flexibility. Among other things, it can help identify the determinant a program should address. For example, a working group wanting to serve a community by decreasing summertime delinquency (the outcome) among disadvantaged youth (the target group) decides to begin its task by identifying a tractable cause (for some, a more familiar term is *leverage*) of the problem—that is, a determinant. The group may determine that the proper determinant is involvement of youth in legitimate social, recreational activities. It would next specify an intervention, such as summer camp, that is expected to activate the determinant, in turn alleviating the delinquency.

Four steps from the development facilitation strategy remain to be discussed. Backward reasoning will provide the basis for their explication.

1. *Identifying the Problem.* The first requirement of development facilitation is to identify the problem a program aims to address. This is a relatively easy assignment for members of the working group (or interviewees). Still, if needs assessment or formative research data are available before the scheduled working

group meeting (or interviews), these data can benefit stakeholders immensely with their systematic exploration of community needs. With such data at hand, the stakeholders' decision should be an informed one.

2. Identifying a Target Population. The next requirement of development facilitation is to specify a target population or group of the program. Evaluators can foster a reasoned choice of target by asking the working group (or interviewees) to consider two things: the requirements or preferences of the funding agency and community needs that may have been pinpointed by needs assessment. Funding agencies frequently note, in the application materials for offered grants, that specified populations must be served by a funded program. Obviously, agency-issued guidelines (general or specific) for naming the target group must be heeded.

Furthermore, it is very desirable that programs serve those people most in need or at highest risk; this must be the top priority in most cases. Exactly who is neediest or most vulnerable, however, is likely to vary from community to community. With HIV transmission, for example, some communities' high-risk population is men who have sex with other men, whereas other communities' most vulnerable members are people who inject drugs (or migrant workers, or the homeless, or sex workers). Those working group members with long experience working directly with clients are good sources of information about community groups that have unmet needs. (Most working groups include at least a few such members.) Applying needs assessment or formative research—or perhaps gleaning existing data, such as epidemiological reports or vital statistics—also generates such information.

3. Identifying Final Goals and Measurable Outcomes. The next step in the development facilitation sequence is to finalize the program goals and establish measurable outcomes to prove their attainment, in light of the specified problem and target group. General directions, and perhaps a list of goals, often come from a funding agency or foundation, appearing in the grant application or call for proposals. Also, stakeholders usually have firm ideas about what goals a program should pursue. In moving stakeholders toward agreement on a program's goals and outcomes, evaluators should foster discussion of several issues.

The first is desirability versus plausibility. Stakeholders tend to formulate goals that reflect very high expectations; these are outstanding aims whose desirability cannot be contested. Although glorious statements of program goals can inspire people and help build coalitions, they can also be unrealistic. A program with goals that are desirable but not plausible is destined to fall short of the mark. From the program evaluation viewpoint, programs need

practical goals—goals that can be reached with the resources available and that have some connection to existing knowledge and experience. Program designers who wanted to reduce drug use in a community largely by distributing “Say No to Drugs” buttons, for instance, would find scarcely any data indicating that a button distribution program is leverage enough to curtail drug use. For this program, it would be unrealistic to establish a goal of reducing drug use in the community. The evaluator does a favor to stakeholders when he or she emphasizes that the statement of program goals must exhibit plausibility and practicability in terms of available resources, proposed interventions, and the nature of the community problem. For example, suppose a media campaign against racism is proposed that would center on statements, published in newspapers and broadcast by radio, iterating the need to work together. The stakeholders in this antiracism program might be inclined to say that the goal of their campaign is less racism. This would certainly be a noble goal, one attractive to many stakeholders and others, but racism is a deeply rooted, complicated problem and has been for many years. It is questionable, to say the least, that racism could be solved by a brief media campaign. Should this program ever be evaluated based on such a grand goal, it would have to be deemed a failure. To prevent this, the evaluator works with stakeholders to arrive at practical goals for the program: perhaps, in this case, enhancing people’s awareness of the problem of racism or building support for long-term strategies to reduce it.

Goals themselves can be long-term or short-term. Short-term goals are attainable in a few months to a year, and most can be achieved. Long-term goals are the ultimate aims for which a program strives. As *ultimate* goals, they typically require a great deal of time to achieve; furthermore, their achievement is contingent on many factors. A short-term goal for a homeless program might be to shelter homeless people for a certain period. Its long-term goal could be to help these same individuals find work, attend job training, or obtain more permanent housing. Assisting homeless people in finding and keeping jobs is a much more challenging task, however, than locating a bed for the night. Many months could go into addressing job readiness alone because issues such as inadequate job skills or unreliability due to mental illness or substance abuse will affect the outcome. It would clearly take a long time for a program to manage these issues and achieve its long-term goals.

Nevertheless, long-term goals are very important to stakeholders in many situations; such goals are the reasons an organization or agency is entrusted with funds. The point is that the working group must identify short-term goals as well as long-term ones. Short-term goals easily provide the “measuring sticks” with which program staff can see their successes and be motivated to

press on with their work. The evaluator should try to see that a working group finalizes both short-term and long-term goals. The desirability/plausibility and short-term/long-term issues have a certain relationship. To attract funding, program stakeholders may believe it necessary to set extremely desirable—but simultaneously unrealistic and impracticable—goals. Indeed, it may be possible for a working group to embrace such goals in the long term, while setting more practical goals in the short term. In the example of the media campaign, for instance, reducing racism could indeed be the long-term goal of the messages, while enhancing awareness of racism and the supports available to deal with racism could be short-term goals.

An evaluator participating in a working group may soon find that stakeholders sometimes confuse goals and objectives with action steps. The evaluator should make the differences clear. *Action steps* are activities that create or strengthen an implementation system. Action steps, then, relate to the elements of the action model. Examples of action steps are “to hire three additional outreach workers and a program coordinator” or “to build an information technology system that supports sound fiscal management.” *Objectives* are the achievements of the implementation on its way to reaching its final goals. In many cases, objectives are given in terms of action steps, meaning that statements such as “to hire three additional clinicians in the first quarter of the project” or “to provide treatment to 50% of clients in 3 months” should not be regarded as program goals.

In addition, the evaluator may need to reiterate to stakeholders the distinction between goals and outcomes. Ideal statements of program *goals* are concrete and concise, indicating the program’s purpose and conveying key stakeholders’ interests and concerns. Most, however, speak generally and in the abstract. Thus, it is even more necessary for the working group to go further, not stopping with statements of program goals but also—based upon those statements—finding a way to express the goals in clearly defined and measurable terms, which are called *outcomes*.

Clearly defined program outcomes are fundamental to the operation of any program. Outcomes provide tangible yardsticks for measuring the accountability variables that funding agencies and/or the public may scrutinize. Could a substance abuse program satisfy those who hold it accountable by telling them that the program really did assist drug abusers in reaching a meaningful reduction in drug use and achieving social and economic functioning? Probably not, because these terms are not measurable. What exactly do the words “meaningful reduction” mean? Is cutting back from drug use three times daily to twice daily “meaningful”? Similarly, “social and economic functioning” is not a precise, measurable concept. To be useful for accountability

purposes, a program goal must be rendered measurable—in which case it has become an outcome.

Consistency among program goals can be as important as clarity. An intervention program often has multiple goals. One task for the working group is to ensure that these goals (and corresponding outcomes) are not incompatible. The program evaluator must be prepared to point out to stakeholders any goals or outcomes that are inconsistent or even mutually exclusive. For example, suppose that a family court proclaims two goals for its handling of child abuse cases: providing abused children with security and support and keeping families intact. At times, these two goals may simply be incompatible. Pursuing the goal of keeping families intact might require returning children to abusers, thereby compromising their safety. Conversely, pursuing the goal of providing security and support to children might require sheltering them apart from their biological parents, in which case the family is no longer intact.

4. *Identifying Determinants That Are Likely to Change Outcomes.* Tied into choosing achievable goals is understanding which determinants are at work in the problem being addressed. Again, determinants are leverages or forces—the causes of a problem believed to be linked directly to the production of outcomes. Some stakeholders may substitute terms like *intermediate outcomes*, *root of the problem*, *mediator*, *social and psychological facilitator*, or *social and psychological barrier* for *determinant*. Evaluators need to figure out which term is being used by stakeholders and communicate with them accordingly.

Once activated or changed, determinants can change outcomes. Evaluators can show their working groups which determinants must be their focus if a program is to attain its goals; this is the next step in completing the program scope. The causes of problems can be many. Spouse abuse can arise out of husbands' low self-esteem, wives' lack of economic power, husbands' responses to stress, the patriarchal society, husbands' ignorance of the law regarding spouse abuse, or any number of other factors. Which of these determinants can or should an intervention program address? Consulting with a content expert or performing a literature review can educate working group members about determinants that have potential to create change. Because a program is usually constrained by its resources, the working group needs to identify one, or at most a few, major determinants in line with the mission and expertise of the implementing organization. It might be appropriate, to continue with our example, for an intervention program against spouse abuse to limit itself to the determinants of low self-esteem and unskilled response to stress.

The selection of determinants has a direct bearing on the selection of interventions. A program intended to undo an increase in youth crime, for example,

will be able to address only some of the potential factors that drive youth delinquency: insufficient parental supervision, peer pressure, poor school performance, drug abuse, inadequate recreational opportunities, child abuse, and lack of positive role models. Perhaps designers of the program decide to focus on one cause, peer pressure, as the determinant. Their choice means that interventions built into the program will also focus on peer pressure.

Determinants and Types of Action Model/Change Model Schemas

A program's determinants are identified according to the type of program theory on which the program is based. Programs based on formal theory (see Chapter 3) can usually find within that theory some statement of appropriate determinants. For example, a program built on the theory of planned behavior (Ajzen & Fishbein, 1980) will draw from that theory a wealth of information about *intention to act* as a fundamental determinant in behavioral change.

Programs built on stakeholder theory, by contrast, receive no such guidance, leaving stakeholders responsible for examining the assumptions underlying their choice of determinants. The evaluator can help clarify the stakeholder theory, if necessary, by making implicit assumptions explicit. Steps in this facilitating process include providing examples of determinants to acquaint stakeholders with the concept; asking stakeholders to name the determinants they believe will most affect program outcomes; listing these major determinants; evaluating these determinants in light of time and resource constraints on the program; and, finally, asking stakeholders which determinants they desire to *and can afford to* select as their program's major focus. When it has been particularly difficult for interviewees or working group members to latch onto the concept of the determinant, the evaluator can try substituting the terms *intermediate outcome*, *cause of the problems*, *social and psychological barrier*, or *social and psychological facilitator* for *determinant*. Then, probing questions from the evaluator might take such forms as "Can you name intermediate outcomes that the intervention needs to reach first before attaining the ultimate outcomes?" "Can you name crucial social or psychological barriers that prevent clients from achieving program goals (which the program would seek to remove)?" and "Can you name crucial social and psychological facilitators helpful to clients who are trying to achieve program goals (which the program would seek to provide or enhance)?" The listing of potential facilitators and barriers that should emerge will be the equivalent of a listing of major determinants.

A program in Taiwan offers one effective illustration of the dynamic when a working group identifies the cause of a problem. Such a group was formed there in the 1980s to respond to increased suicides by police officers during a transition from a one-party political system to a multiparty system. In a meeting, some members of the group, including psychologists, argued that these suicides were mainly attributable to police officers' lack of stress management skills and that stress was intensifying because mass demonstrations had continued in the wake of abolishment of martial law. These members advocated conceptualizing the suicides as a problem of individual adjustment. Other members of the working group, however, including sociologists and program evaluators, suggested that other influences should be considered. One of these was the sudden abeyance of social norms following the demise of martial law. Under martial law, there had been no salient structural distinctions between the ruling political party—the Komington—and the government. The Komington was the government, and vice versa. Police officers belonged to the Komington's law enforcement arm; their authority was legitimate. As martial law crumbled, however, that authority weakened, and the public now regarded police officers as puppets of the Komington. The police were ridiculed and humiliated in public and in the media. This abrupt identity crisis might have something to do with the increase in suicide by officers, the second faction countered. The decision eventually made by this working group—in its entirety—was to place their program on the individual level rather than the structural level. By identifying personal turmoil as the determinant, the working group felt that it would subsequently be straightforward to design and implement a counseling program for police officers. The working group also, however, notified decision makers of the structural problem and recommended that they address it. This exchange among stakeholders and evaluators in Taiwan shows how alternative views can be brought to stakeholders' attention with good result. At the same time, it demonstrates the value of barring evaluators from pursuing personal agendas or substituting their own values for stakeholders' values. In short, an evaluator's role is to ensure that stakeholders make *broadly* informed decisions.

Choosing Interventions/Treatments That Affect the Determinant

When determinants have been identified adequately for the purposes of the program, the evaluator's work shifts to facilitating the working group's selection of an intervention that can activate those determinants. There are usually a number of intervention options. For example, if a program scope posits that

truancy is a major determinant of burgeoning youth crime, and if truancy therefore will receive the program's focus, there are many ways to try to change this determinant. Schools might provide counseling to help students with academic or other school difficulties. They might develop new curricula to appeal especially to troubled youths. A policy might be enacted that fines parents when their children are absent; sends school administrators to immediately visit any student who has not shown up for class; or authorizes police to write citations to students not in school during school hours, requiring both the student and parent to appear in court. Pinning down the best intervention involves four criteria, which the evaluator should share with other members of the working group:

1. *Mission and philosophy of the organization.* The intervention selected by the working group must be appropriate to the implementing organization's mission and philosophy. Conflict between the intervention strategy and the implementing organization's values creates stress and can interfere with implementation of the program.

2. *Budget and personnel restraints.* The chosen intervention must reasonably reflect the budget and expertise of the implementing organization. No organization can effectively execute an intervention far beyond its means, nor can it responsibly agree to an intervention requiring personnel it does not have.

3. *Theoretical justification.* An intervention needs sound theoretical justification in order to be effective. After all, when theoretical justification is scant or weak, it is scant or weak for good reason. As we have seen, a program founded on scientific theory finds well-reasoned determinants in that theoretical ground, and stakeholder theory-based programs can justify a choice of determinants by citing other programs, literature, or common sense. It is possible to justify a job-training program (which is an intervention) by noting that it enhances clients' job skills (which is the determinant) and by noting further that better job skills can lead to employment or better-paid employment (which is the program goal). It is possible to do this because the relationships among job training, acquired job skills, and employment comprise a well-recognized common experience that makes sense to most of us. It is less possible to convincingly justify "say no" buttons (an intervention) as an influence on perceptions of drug use (the determinant) capable of discouraging drug use in a community (the program goal), because our experience tells us the intervention is too weak to produce such a characteristically hard-won change.

4. *Base of evidence.* Other things being equal, intervention strategies that are supported by empirical research should be preferred over intervention strategies

without such support. This principle is not meant to discourage innovation, but the existence of empirical evidence, even indirectly linked empirical evidence, does provide additional confidence in an intervention.

THE RELEVANCY TESTING APPROACH: A PART OF THE TROUBLESHOOTING STRATEGY

In some situations, program designers have completed a program scope but wonder whether the proposed change model it includes is relevant to the problem as it is experienced by target group members in their world. A program evaluator might be called on to carry out empirical checks—field-testing—on the program scope before deploying it to create a program plan. The *relevancy testing approach* can meet the needs of program designers who are set this task. Relevancy testing comprises a reality check. It is a rapid appraisal, from the field, of the degree of fit between the problem faced by the target groups and the selection of interventions, determinants, and goals for a program. It is a very useful approach for identifying potential weaknesses and improving the quality of a program scope. The stakeholders engaged in relevancy testing are most often the implementers and clients of the program. Like formative research, relevancy testing is flexible as to the research methods involved (focus groups, interviews, and surveys are options). The program evaluator conducting relevancy testing is likely to find three questions very important to her or his progress:

1. *Are the goals stated in the program scope appropriate and reasonably relevant to the identified problem and target group?* In other words, is it realistic to envision the targeted individuals achieving the established goals? Sometimes, the original program scope includes goals beyond the scope of the program's interventions or ill-fitted to clients' problems and needs. Weeding out such inconsistencies begins with asking clients and implementers simply to comment on the program scope. As an example, a perinatal care program prepared a program scope proposing to employ a midwife to care for 300 pregnant women, and deliver their babies, in 1 year. The evaluator brought together a group of midwives and asked for comments. The group's immediate response, based on its extensive experience, was that meeting this goal was impossible. Without resources above and beyond those available to the program, the group felt no more than 100 women could be served in a year.

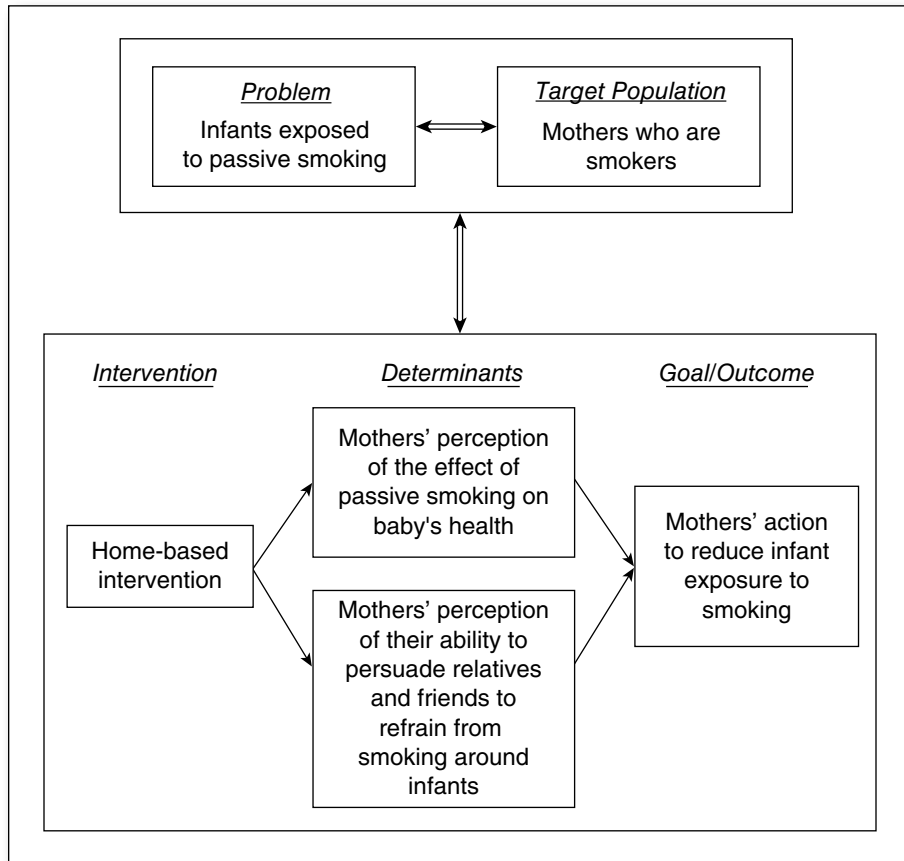
2. *Does the program scope name a determinant likely to have an impact on the target population?* A program scope assumes that a particular determinant

causes a problem affecting a target population, and it proposes an intervention to alter the determinant, creating desirable outcomes. Those on a program's front lines—people such as clients and implementers—can help the evaluator understand whether a determinant vital to the program scope is sufficiently relevant to the proposed target population. Teen pregnancy prevention programs provide an example. Such programs may assume that adolescent girls become pregnant because they are unskilled in using birth control devices. Making that assumption, such programs set out to teach the girls about the devices, both to make them comfortable with using them and to prevent frustrating, unsuccessful attempts at using them. The program's two determinants are, then, lack of skill and frustration. To test the accuracy of these determinants, the evaluator invites teenage girls to participate in focus groups. If discussions suggest that the girls in fact know a lot about using birth control, the evaluator will tell program planners that the selected determinants may not be appropriate for this target population.

3. *Will the outlined intervention be reasonably acceptable to the target group?* An intervention that is well received by one group can be offensive, or completely irrelevant, to another. The evaluator engaged to field test a program scope must ensure that the proposed intervention or treatment will be both relevant and inoffensive to the particular population being targeted. One major issue is cultural sensitivity. Some societies are more conservative than most Western societies. A lesson on birth control that includes the demonstration of condom use before a group of young women will not be helpful if the audience finds this insulting or painfully embarrassing. Making inquiries about the cultural sensitivity of proposed interventions is one task for the program evaluator. Another is determining whether clients will understand the implicit or explicit messages of an intervention. This determination involves both clients and implementers. Synthesizing the comments of these two parties, the evaluator learns whether the language of the message is understood by clients and is respectful of their culture, and the evaluator learns whether implementers feel confident that they can effectively communicate to clients the proposed curriculum.

Research Example of Relevancy Testing

In Chapter 3, a home-based intervention program to reduce passive smoking by infants (Strecher et al., 1989) was used as an example, and we return to it here to illustrate relevancy testing. This program's scope is represented graphically in Figure 4.2.

Figure 4.2 Scope of a Program to Reduce Infants' Exposure to Smoking

SOURCE: Adapted from Strecher et al. (1989).

Strecher and colleagues (1989) tested the relevancy of this scope in the field by recruiting 104 mothers of infants; 40 were recruited at clinics for face-to-face interviews, and 64 were identified via county birth records available to the researchers and contacted for a telephone survey. The researchers obtained information that answered the three questions above and clarified the assumptions underlying the program scope. First, for example, they asked whether the program's stated goals were relevant to the target population. Program planners had raised the notion of urging mothers to stop smoking altogether, but the researchers wondered whether a more modest aim (providing a smoke-free

environment for the infants) was more realistic. They queried the mothers about their interest in giving up cigarettes; the resulting data showed that most would not be inclined to try to quit smoking during this period in which they already faced the many stresses of caring for a new baby.

The researchers also asked whether the named determinants, outcome expectation and efficacy expectation, were likely to have an impact on the mothers. Outcome expectation comprises one's belief about whether a given behavior will lead to a given outcome; efficacy evaluation comprises one's belief in one's own capability to perform the behavior that leads to the outcome. In the example, the obtained data showed that mothers had general knowledge of the effects on babies of active smoking by adults: respiratory problems. However, according to the findings, many mothers did not understand the nature of passive smoking. For example, they believed that a baby inhaled significant amounts of cigarette smoke only when the smoke was blown directly into the baby's face. Few realized that a baby also inhales significant amounts of smoke when it is across a room from a burning cigarette or when it enters a room where people have been smoking earlier. Findings like these informed the researchers about the outcome expectation determinant. Other findings informed them about the efficacy expectation determinant. In general, the researchers found low levels of perceived efficacy among the new mothers, reflecting a lack of confidence in their ability to persuade husbands, other relatives, friends, and caregivers to refrain from smoking around the infants.

Strecher and colleagues' (1989) relevancy test prompted fine-tuning of the program scope in the following areas: An emphasis was placed on teaching mothers what passive smoking actually is; the intervention was refocused on the immediate, specific outcomes of an infant's exposure to secondhand cigarette smoke; and the program aim was reoriented toward helping mothers maintain smoke-free environments for their babies. The issue of smoking cessation was discussed only if a mother or other family member expressed interest in quitting.

Moving From Program Scope to Action Plan

In this chapter, the importance of the program scope has been demonstrated, including its position as the basis for program plan development. In addition, we have reviewed ways to facilitate stakeholders' development of their program's scope. Once a sound program scope has been adopted, stakeholders proceed to devise the action plan component of the overall program plan. Program evaluation is a valuable facilitator of this work as well, and this is the topic of Chapter 5.

QUESTIONS FOR REFLECTION

1. How are program scope and action plan different? Explain by using real-world or hypothetical examples.
2. If a program is planned without a clear program scope first being developed, what may be the results? Why?
3. What are the components of a program scope? How does program scope relate to the change model discussed in Chapter 3?
4. What are the purposes of needs assessment and formative research? Is it essential that a needs assessment and formative research be conducted every time before an intervention program is developed? Why or why not?
5. How are unmet needs in a community usually identified? Give real-life examples, from research or organizations you have worked with, that illustrate where one could obtain information about a target population's unmet needs.
6. How does formative research allow an evaluator to explore an issue more deeply?
7. What are the advantages and disadvantages of conducting focus groups?
8. Consider the example in the text of the Hawaiian program developed to teach primary school students about preventing skin cancer. Would a needs assessment be appropriate for this program? Defend your answer.
9. How might the conceptualization facilitation approach be put into practice? Give an example.
10. Identify the conditions that favor the use of the working group or intensive interview format.
11. In your opinion, can the terms *determinants* and *identifying causes of the problem* be used interchangeably. Why or why not?
12. Why is it important to identify a target population for a program? In what ways can an evaluator and stakeholders do this?
13. Distinguish between desirability versus plausibility in terms of identifying program goals. Give an example.
14. Why should evaluators focus on evaluating plausible goals? Explain.
15. What is a relevancy test? Give an example and illustrate the relevancy test's importance to the program in your example.

CHAPTER 5



Helping Stakeholders Clarify a Program Plan

Action Plan

Program evaluators who are asked to help with developing the program may be asked to help formulate the program scope and action plan of a program plan. In this chapter, evaluators' facilitation strategies and techniques are presented, as they can be used with start-up intervention programs. However, the discussion may also prove useful for established programs whose stakeholders decide to fine-tune or revise an existing action plan. As Chapter 4 sought to make clear, a sound action plan is one that has been guided by a program scope. So, again, the evaluator asked to facilitate the production of an action plan in the absence of a program scope will need to approach the stakeholders about the importance of adopting a scope before moving on.

THE ACTION MODEL FRAMEWORK AND THE ACTION PLAN

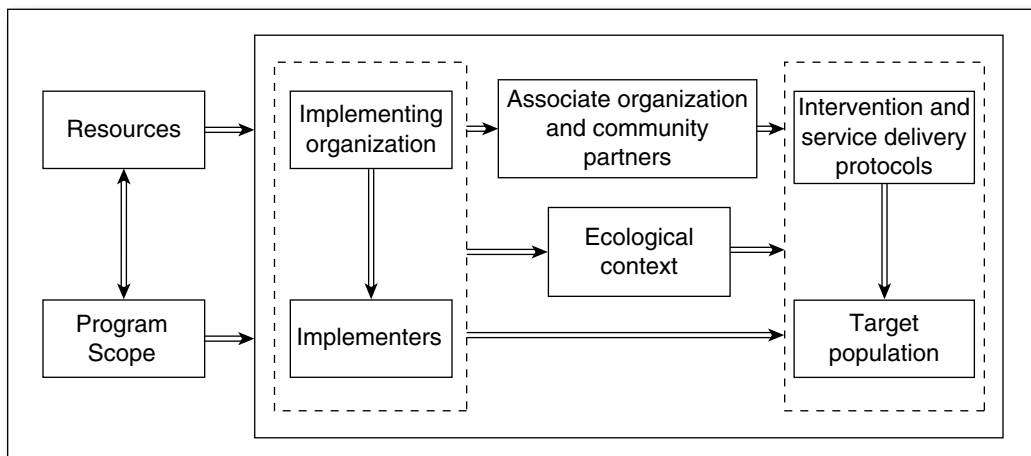
From the stakeholders' viewpoint, action plans are blueprints for the activities prescribed by program scopes. Early on, action plans guide the organization of program activities and the allocation of resources. Later, they stipulate the program staff's day-to-day operations and coordinate disparate personnel units. The quality of a program's action plan affects the quality of its implementation and, eventually, the degree of its effectiveness. High-quality action plans come most readily from clear and realistic program scopes. A program scope that never quite says how to reach and screen the target group, for example, complicates

the work of program managers as they try to determine the kind of implementers to recruit and the kind of training the implementers need.

Just as the program scope belongs to a program's stakeholders, the action plan is theirs, too. It is the stakeholders' duty to develop the plan for their program. However, when a program's complexity requires it, stakeholders seek help from experts, like program evaluators, to ensure that the action plan is sound and feasible. An evaluator invited to facilitate the action plan's development can draw on a general knowledge of action models (which is part of the action model/change model schema). The number of activities called for in most programs can be dizzying, even to stakeholders. They pose a challenge to evaluators as well, who may need to help stakeholders conceptualize all of the various activities within a meaningful scheme so that implementation can be successfully managed. As a guide to that scheme, this chapter offers a general conceptual framework of an action model, the *action model framework*. This framework, which is represented in Figure 5.1, is a conceptualization of a generic action plan, useful in developing any number of specific, situational action plans.

On the left of Figure 5.1 sits the program scope. Its role is to direct the development of the action plan, which includes six components, often considered in a standard sequential order; that is, certain components must be devised before others. Implementing an action plan requires a capable, committed implementing organization. It is up to the implementing organization to find and train responsible implementers, who must have ample skill and commitment. The

Figure 5.1 Conceptual Framework of an Action Plan



implementing organization (and its implementers) also must connect to associate organizations and community partners so that services may be delivered efficiently, and it must generate interpersonal and community support for the program. Yet another job for the implementing organization is to reach clients from the target group and motivate them to join the program. Only with all of these components in place can implementation of the intervention protocol be launched and services delivered to clients.

In spite of their sequential order, all six components are interconnected. When new information alters a component, the other components often change as well. For example, stakeholders may decide to include additional target groups in the program. This move immediately affects the implementing organization's need to coordinate activities, the intervention activities themselves, the content of implementation protocols, and other components. The relationship between program scope and action plan need not always be unidirectional, either. On occasion, in the process of developing the action plan, stakeholders become aware of problems or weaknesses in the program scope and revise it accordingly. Figure 5.1 illustrates this feedback process within the action model framework.

The action model framework is helpful whether the evaluator is facilitating (a) the development of a new action plan, (b) the clarification and strengthening of an existing action plan, or (c) communication about the action plan. It is also useful later in the process in terms of (d) preparation for the formal evaluation of the program's implementation.

Developing a New Action Plan. Program designers initiating an action plan depend on input from program directors, implementers, and other key stakeholders. The action model framework is a means of systematizing the distinct pieces of this collective effort. It helps to ensure that no important issue is left out and that no gaps are permitted to compromise the quality of the action plan.

Clarifying and Strengthening an Existing Action Plan. When stakeholders question the completeness or soundness of a action plan they have devised, the action model framework is a good tool evaluators can use to get stakeholders working together again. This purpose for the framework is at times very similar to the next, facilitating communication.

Communicating About the Action Plan for Future Evaluation Design. The action model framework shows how each of many program activities ties in with the others. Larger programs may include hosts of activities, which the framework allows to be ordered into categories. As stakeholders begin to see the proverbial "method to the madness," they become better able to discuss the action plan with each other and their various constituents.

Planning Ahead. It is generally a given that, once underway, a program will need to be formally evaluated. Usually a process evaluation is requested to research how closely the actual implementation of a program matches the stakeholders' intentions. The action model framework can make clear precisely what their intentions were, a valuable aid to the evaluator. Similarly, when a program requests outcome evaluation, the action model framework can answer the evaluator's questions about how interventions were to be delivered or how the target group was to be reached. Knowledge of such information is essential for designing an outcome evaluation.

STRATEGIES FOR DEVELOPING ACTION PLANS

Calling on a program evaluator to assist in preparing both the program scope and the action plan—as stakeholders frequently do—is certainly advantageous. Evaluators have the knowledge and skills to ensure consistency and that all crucial elements of an action plan are securely in place. Furthermore, the evaluation strategies and approaches for developing action plans meld with those for drafting program scopes, and preparation of the two is, for the practical program evaluator, essentially a unified project.

The strategies and approaches for assisting stakeholders with action plans presented here, with examples of their application, are not mutually exclusive. It is sometimes beneficial to include more than one tactic during development of the plan. The background information strategy can provide stakeholders with general information about communities and clients so they can put their action plans in motion; the formative research evaluation approach fits this purpose well. The development facilitation strategy can be ideal to use with stakeholders who feel they need evaluators' input to develop the plan for their program or evaluators' support to build consensus about the worth of an action plan. The troubleshooting strategy uses small-scale trials of the program action to iron out potential problems and come up with ways to fine-tune the plan. Selecting among, and then using, these strategies and approaches can be complex; guidelines for the process appear in Chapter 3.

THE FORMATIVE RESEARCH APPROACH (UNDER BACKGROUND INFORMATION PROVISION STRATEGY)

Chapter 4 illustrated how the formative research approach generates meaningful background information for stakeholders working on a program scope. The

same principles and procedures of formative research are valuable for work on an action plan. Simply put, the evaluator will find it very feasible to design and conduct *generic* formative research to generate background information that will be useful in composing both the program scope *and* the action plan. However, certain considerations, discussed below, are especially pertinent when working with stakeholders on an action plan. (See Chapter 4 for the more general principles and procedures of formative research.)

1. *Formulating research questions to inform the action plan.* The action model framework can play a role in determining which questions should be researched to equip program designers or working groups with adequate understanding to write an action plan. Examples of research questions are “What factors discourage target group members from participating in the program?” “What components should be included in the intervention protocols?” “What mode of service delivery will be acceptable to these clients?” “Does the proposed implementing organization have the needed capacities?” and “What training will implementers require?”

2. *Gathering data to answer the research questions.* Flexible research methods are a hallmark of the formative research approach. The key need is to return feedback to the program designers or working group *quickly*. Focus groups, intensive interviews, and surveys all can provide good information. Research participants are typically prospective clients, implementers, and other stakeholders.

Example of Formative Research

Formative research was the mode selected by Gettleman and Winkleby (2000), who set out to learn what might be the best structures and implementation schemes for programs addressing incipient cardiovascular disease (CVD). Low-income women—African-American, Hispanic, and White—constituted one of the largest populations at high risk for CVD, but insufficient information existed on how to reach them in meaningful numbers. Gettleman and Winkleby’s research involved seven focus groups with 51 low-income women in their communities. The focus groups showed that these women preferred receiving health information in “visual” formats as opposed to text-only formats. They also felt most positive about prevention programs that addressed several risk factors, particularly smoking, lack of exercise, and high-fat diets. Two more elements of a prevention program the focus groups thought would

be effective were testimonials from healthy women who described how they embraced heart-healthy behavior and factual commentaries from physicians. The CVD intervention emphasized staying healthy for one's own sake, as well as heart-healthy behaviors and skills and the role of choice in behavior change.

Formative research during this project also asked the women about barriers to and incentives for participation. Three identified barriers were a lack of time, transportation, and child care; most said they would not participate in a CVD intervention program unless there was free child care. As to constraints on time and location, the women suggested intervention programs be held at job sites during lunch hours for those employed outside the home and at community sites for others. The researchers concluded that transportation problems could be eased by offering program activities in places women routinely go with their children, such as public libraries or pediatric clinics. The research identified child care, free meals, and cash or food vouchers as incentives with the potential to encourage participation. All of this information proved valuable in developing interventions tailored to low-income women.

THE CONCEPTUALIZATION FACILITATION APPROACH (UNDER DEVELOPMENT FACILITATION STRATEGY)

As was discussed in Chapter 2, within the development facilitation strategy is an evaluation approach called *conceptualization facilitation*, one that can serve very well to help stakeholders with their action plans. Although we treat conceptualization facilitation as a distinct evaluation approach, using it to develop an action plan involves the same principles and procedures that Chapter 4 presented for helping stakeholders develop program scopes. For instance, the communication formats available for developing scopes (intensive interview, working group) are also used frequently in action plan development, with just a little modification. For example, the membership of the working group or the people to be interviewed may be different in that the group will include implementers. Such modification ensures that both administrators and implementers are well represented during development of the plan; each party brings expertise to the work and holds high stakes in the quality and success of the plan. Working from the program scope, the evaluator can discuss thoroughly with stakeholders the elements of the action model framework and how these relate to an action plan.

Once more, the elements, or components, of the action model framework are the target population, the intervention and implementation protocols, the program implementers, the implementing organization, peer organizations and

community partners, and the ecological context. The evaluator's first task when applying the conceptualization facilitation approach is to assess the importance of these six components to the proposed program. Four of the components are considered crucial in designing any action plan; the importance of two others (peers/partners and ecological context) varies from program to program. In general, larger programs should assign relatively more weight to an implementing organization's capacity and to establishing linkages with peer organizations. Furthermore, a program acknowledged to be less than appealing to its prospective neighbors—such as a homeless shelter or halfway house—should give added weight to contextual supports for the program's implementation. When stakeholders are confident that a given component is irrelevant to their program, it may safely be excluded from the action plan. However, the reason for such a decision should be documented, and all stakeholders should made familiar with it and, ideally, give it consensus support.

While helping stakeholders with an action plan, program evaluators typically work to endow the six components of the action model framework with certain qualities. These have been touched on earlier in the book; what follows below is a step-by-step description of the six-part process of helping stakeholders produce an effective action plan. Evaluators can apply a working group or intensive interview format. The following are discussed in the working group format; readers can easily apply the same principles and tactics to the intensive interview format.

1. Implementing Organization: Assess, Enhance, and Ensure Its Capacity

To select an implementing organization and, in the process, to measure that organization's capacity-building needs, a working group should consider three main factors: technical expertise, cultural competence, and manpower and other resources.

Technical Expertise. The working group needs to discuss what technical expertise potential implementing organizations need to have. Which organization is best equipped to deliver a particular intervention? In cases in which a group of implementing organizations has already been selected, the working group can recommend criteria for program directors and other key decision makers to use in their evaluation of the technical proficiency of the organizations.

Cultural Competence. The evaluators can remind the working group of the importance of specifying what kind of cultural background and experience the

implementing organizations need to have in order to recruit implementers who will communicate well with their clients or to be trusted by the target population. The working group can also indicate what kind of training the implementing organizations need to provide to develop and ensure the cultural competence of their implementers.

Manpower and Other Resources. The working group needs to work out a plan to select implementing organizations that can commit sufficient manpower and resources to the program. Implementing organizations that are overloaded with existing projects should not be asked to implement the new program. This can happen when government agencies are tasked with starting new programs by administrative order. The implementing organization often receives the order with no accompanying budget or staff increase. Busy with, or even overloaded by, existing duties, the agency's staff may feel little incentive to take on yet another program.

2. Intervention and Service Delivery Protocols: Delineate Service Content and Delivery Procedures

The general nature of the interventions to be conducted is laid out in the program scope. In developing an action plan, however, the working group needs to go much further. First, it should specify in detailed terms the services to be provided by the program—the program's *intervention protocol*. The intervention protocol is a description of the content, curriculum, intensity, and duration of the intervention services or activities to be provided to the target group. The working group should strive to note every detail. Next, it must explain just as thoroughly the procedures involved in, and also the setting for, the delivery of these services—the program's *service delivery protocol*. No intervention can be carried out precisely in the field without complete intervention and service delivery protocols.

For example, in developing the intervention protocol, the working group of an HIV-prevention program might specify that a group counseling intervention take place over three weekly sessions, each 2 hours long. The working group would then elaborate on these sessions. For example, the first session should be an "icebreaker," with the purpose of freeing clients of some of the reluctance they might feel about speaking openly of HIV risk; the second should be a discussion of barriers to safer sex practices and how to remove these barriers; and the third should teach information and skills clients need to practice safer sex. The working group must arrange for the creation of a curriculum for each session in order to provide implementers with clear guidance as to topics and activities; the curriculum should also include any available advice for making

each discussion session a success. In developing the service delivery protocol, the working group will specify where the sessions will be held and how. An evaluator facilitating the work of drawing up protocols should be certain the program designer or working group members are familiar with potential modes of service delivery and potential service settings. *Modes of delivery* include the following:

- *One-on-one interaction*: An intervention delivered by an implementer to an individual client (one client at a time), such as individual counseling. This is labor-intensive and therefore costly, but it is possibly one of the most effective modes of delivery.
- *Support group*: An intervention via a group process of mutual understanding and support, such as Alcoholics Anonymous. A group of several clients plus a therapist/facilitator meets regularly, opening minds to the notion of change, mutually encouraging and accepting change, facilitating change, and sustaining change. (One-on-one interaction and support groups are, of course, popular modes of delivering treatments as well as interventions.)
- *Intervention classroom*: Information or demonstration (e.g., a brief exercise routine) delivered to target group members by a presenter, often followed by question/answer time. This is inexpensive to provide but difficult to individualize.
- *Documents/literature*: Intervention message published in a brochure, pamphlet, flier, or similar document and mailed to target group members or distributed for pickup in public areas. This mode has the potential to reach a range of people at relatively low cost, but there is a danger that information-weary readers will ignore or take lightly the published message.
- *Telephoning/Web posting*: Calls from implementers to target group members to deliver an intervention, or intervention messages posted on websites thought to be frequented by target group members. The effectiveness of this mode of delivery is constrained by clients' and implementers' access to technology.
- *Mass media*: Conveying the intervention message via television and radio, newspapers and magazines, hotlines, and so on. This passive delivery method offers no assurance that the target group will encounter the message.

An implementation protocol must also specify the desired service setting. Informed decisions can be made only when the working group or program designer is aware of the range of possible intervention settings. These include the following:

- *Office/clinic/hospital*: This setting often belonging to the implementing organization and is under its control. This offers a logistical advantage *if*

target group members have transportation to the facility. The professional surroundings may impress clients, or they may feel too formal for the comfortable disclosure of personal problems.

- *School/community center/club facility*: This setting is often conveniently located, but facility directors may balk at providing space for use by certain populations (e.g., drug users, the homeless).
- *Public area* (e.g., street, park, playground, and so on): These settings are open to all. They are popular with outreach workers for making initial contact with clients, but services are often not deliverable on the spot. There is only a slim chance of having more than a brief interaction with busy people who are going somewhere, supervising children, engaged in sport or exercise, etc., so the time is sufficient only for information sharing, not for conducting a real intervention.
- *Store/shop interior*: Examples include a Laundromat, bar, hair salon/barbershop, pool hall, drugstore, bookstore, and so on. Such places are popular with outreach workers seeking target group members. The prospective client's attention may be better focused than in outdoor areas. However, the proprietor's approval is required and may be withheld if the business owner fears an adverse impact on the business.
- *Private home*: The home may be that of the implementer, a volunteer, or a client. Serving food and beverages can create a cheerful, casual atmosphere in which a sense of security and relaxation prompts open sharing of experiences.

When the program being planned is of the extended, labor-intensive variety—a mental health or alcoholism treatment program, for example—clients typically complete a series of stages, perhaps in various settings. The implementation protocol then must specify exact procedures for moving clients from one stage to the next: intake, screening, assessment, referral, treatment, revisit, and finally exit. Among these procedures should be safeguards preventing a client from “falling between the cracks” along the way.

3. Program Implementers: Recruit, Train, and Maintain for Competency and Commitment

Program implementers—the people who deliver intervention services—can be professionals or volunteers. As a working group or program designer explores who will serve as implementers, the program evaluator can assist by providing information on means to ensure the quality of their work. Stakeholders

and the evaluator may also need to concern themselves with raising implementers' levels of technical and cultural competence; determining incentives to encourage implementers' commitment; and communicating clear direction, ample instructions, and firm expectations to implementers about their work. Programs planning to employ professionals need to recruit and select the best available. Most intensive treatments and interventions require highly trained individuals, such as certified teachers for education programs, therapists for drug abuse counseling, and social workers for case management. In programs like these, compromising on the qualifications, experience, and commitment of personnel directly affects outcomes. Unqualified volunteers should never be substituted for professionals when services require professional training. But volunteers can accomplish wonders for some programs—and not just in terms of cost savings. Volunteers who have ethnic, social, and economic backgrounds similar to those of a program's target population may be most capable of making connections with hard-to-reach clients, delivering information to them, and escorting them to the intervention setting.

Assuring the quality of implementers' work is achieved using any of three general strategies, according to the preference of the working group or program designer. The evaluator can facilitate decision making in this area as well. The three strategies are *training*, *technical assistance*, and *review*. Training provided by the implementing organization can establish or enhance implementers' skills and cultural competence, enabling them to deliver services effectively. Even well-prepared implementers, however, will occasionally meet with difficulty. For those times, the implementing organization needs some mechanism to assist implementers, particularly when the program being implemented is groundbreaking or very large. Finally, a crucial part of quality assurance occurs as supervisors' periodically review implementers' work. At times, for fairly obvious reasons, implementers may hesitate to share problems and mistakes with supervisors. If this becomes a concern, perhaps the review of work can be conducted by peers. In fact, peer review often prompts valuable sharing of experiences in handling problems, frequently revealing innovative means for addressing them.

4. Associate Organizations/Community Partners: Establish Collaborative Relationships

An implementing organization may need to create meaningful working relationships with various peer organizations. It is up to the working group or program designer to identify both the organizations of interest *and* the strategies that are likely to launch such relationships. (This step is, as previously noted,

more important to the success of large-scale programs than small-scale ones.) The evaluator should lead the working group in a consideration of the four main types of associate organizations with which the implementing organization might partner: core organizations, related organizations, auxiliary service organizations, and collaborating organizations.

Core organizations are those with which the implementing organization must work—very closely and efficiently—if its program is to be implemented. An education program targeting teenage dropouts obviously needs a strong working relationship with schools, for the schools know both who has dropped out and, often, why. Furthermore, for many intervention programs aimed at minors, parents must be considered a core organization, in that their formal consent is required before their children can legally participate. For example, sex education programs for high school students typically require each enrollee to have a parental consent form on file.

Related organizations are those that have the power to inadvertently interfere with program implementation if they are unaware a new program is underway. For example, a program whose outreach workers will frequent districts known for drug trafficking or the sex trade would be wise to contact police departments in those areas first. Simple notification of this sort has saved many a program from difficulty. Sometimes, a new program must not only notify a related organization but actually obtain that organization's permission to begin implementation.

Auxiliary service organizations can meet clients' needs beyond those the implementing organization addresses. For example, clients in drug treatment programs may also need shelter, food, work, and education. An implementing organization should establish an excellent referral network involving public and/or private social service agencies whenever it observes that its clients could benefit from additional interventions.

Finally, *collaborating organizations* are those whose duties include coordinating services very similar to those the new program plans to offer. For instance, a community-based organization working to implement a new antismoking program could benefit from working closely with a state health department, whose deeper pockets may allow it to assist the program with training implementers, updating technology, and bringing additional resources to bear.

5. Ecological Context: Seek the Support of the Environment

Like peer organizations/community partners, ecological context is a component of the action model with which only some programs need to be concerned.

Ecological context is addressed in an action plan at the discretion of the working group or program designer. If a decision is made to include ecological context in the program theory or plan, both microlevel and macrolevel contextual support may be important.

To gain *microlevel contextual support*, working group members should ask whether the success of an intervention will depend significantly on support clients receive from their social milieus or adjustments to clients' physical surroundings. An education intervention, for example, will usually experience a better outcome when the parents of the targeted students are strongly supportive. Depending on the nature of its program, a working group may need to consider the roles of clients' spouses/significant others, relatives, friends, neighbors, or co-workers. The group may also need to investigate target clients' physical surroundings: homes, schools, workplaces, and neighborhoods. An example illustrates why such considerations are important. Say an educational program is using innovative homework assignments to raise disadvantaged children's reading and math scores. This program could see better outcomes if it provided each student with a desk at home, because altering the physical surroundings in this way might foster sound study habits. Or, think of an antigang program that targets juveniles in at-risk neighborhoods. The intervention message would be strongly underscored by certain changes to the physical surroundings: addition and/or repair of street lights, restoration or removal of rundown structures, prompt erasure of gang-related graffiti on fences and walls ("tagging"), and an overall sprucing up of sidewalks and streets.

Macrolevel contextual support refers to the degree of potential support or opposition an intervention program faces from its community and local institutions. The working group must certainly consider macrolevel matters; if it detects a lack of support (or active opposition) that could hinder implementation of the program, it should incorporate into the action plan some means of securing wider support for it. Some potential strategies for building macrolevel support are media campaigns, consortiums, "summits" with opposition leaders, and community mobilization.

- *Media campaigns*, including television or newspaper ads, draw attention to a problem and show people why a program is needed. Although media campaigns educate the public, they are unlikely to sway institutions or organization leaders who are publicly opposed to a program.
- Joining or launching a *consortium* is another option. A consortium, or coalition of complementary organizations and agencies, exists to support and strengthen its members' agendas. Sometimes, the backing of a consortium helps win support for a program from the ecological context.

- *Summit-style meetings* with leaders of the communities or institutions that oppose a program can also be support builders. In many situations, opposition to a program grows out of a misunderstanding of the program. It is sometimes possible, through exchange of ideas, to agree on a compromise that speeds along the implementation of the controversial program.
- *Community mobilization* is a grassroots and comprehensive strategy. Volunteers are recruited and trained to systematically contact the public, opinion leaders, and officials of organizations and communities to generate support for a program. In a highly mobilized community, former opponents may ultimately become involved in implementing the program.

6. Target Population: Identify, Recruit, Screen, and Serve

As Chapter 4 related, the program scope identifies a target population or group, but the evaluator employing conceptualization facilitation goes further, defining the target population rigorously and specifying strategies to recruit target population members to the program. Managing the target population component of the action plan can be an involved process. The reality of resource constraints usually makes it impractical to target the entire population at risk. Thus, although a target population may be identified in the program scope in general terms, a precise specification or definition is ultimately needed. The program scope of an HIV-prevention program, for example, might cite a target population of “individuals at high risk for HIV.” There are, however, many groups of individuals at high risk of HIV exposure: immigrant workers, men who have sex with men, transgendered people, the homeless, sex workers, and intravenous drug users. Only a very wealthy program might attempt to serve them all. So, the action plan for this prevention effort must include clear eligibility criteria individuals must meet to be included in the target population. Often, these criteria reflect need or deficiency: low income, undereducated, abuse survivor, and so on. Alternately, eligibility criteria may rest on demographic characteristics such as age or residency.

Action plans must be careful to avoid overcoverage or undercoverage of the pool of targeted individuals. *Overcoverage* results when eligibility criteria are too broad and loosely defined, allowing into the program those who do not really need its services. An example of overcoverage would be a government program intended to help small farmers that winds up serving significant numbers of well-to-do farmers with large farms, or even corporate farms, because

its eligibility requirements are too elastic or imprecise. On the other hand, *undercoverage* denies eligibility to numerous qualified members of a target population because the program's eligibility criteria are too restrictive. Overcoverage can waste valuable resources, and undercoverage hamstrings the entire program. The evaluator can assist stakeholders in creating eligibility criteria that avoid both, achieving appropriate coverage.

The target population, once it is defined, learns about the benefits of a program through recruiting. Common recruiting strategies include systematic marketing using radio, newspapers, magazines, pamphlets, and/or other mass media, referred to as a media campaign; referrals by agencies or organizations already serving members of the target population; and outreach. An effective media campaign will home in on places where target group members go about daily activities, including the neighborhood in which they reside. But for many programs, simply communicating to the target population is not enough. Often, outreach workers are ultimately needed to bring clients into programs. One popular outreach strategy engages former program clients as volunteer workers taking on "hard-core" cases or especially high-risk individuals. Former clients know where and how to reach potential new clients, and if they are volunteers, using them also conserves program resources. Initially, though, the program needs to pay for at least some training of volunteers to ensure the quality of their outreach activities.

Target population members reached through media campaigns, referral, or outreach must almost always complete a diagnostic *assessment* as their first step in the program. Diagnostic assessment provides insight into an individual's unique problems, pointing the way to the kind of intervention and social services needed. The assessment is especially important before an individual is admitted to a more intensive type of intervention program (mental health care, substance abuse counseling/rehabilitation, etc.). Perhaps the major purpose of the assessment is to determine whether an intervention by the program will be beneficial for the client. Many health-related programs, especially, depend on careful diagnosis of each interested person to establish the medical necessity of the intervention. (Would-be clients found to not need an intervention after all, but likely to benefit from services beyond the assessing program's scope, can easily be referred to other agencies in most cases.)

Once entered in an intervention program, clients are, unfortunately, likely to drop out again at some point. Excessive barriers to participation and low motivation are two reasons for dropping out of a program. An action plan that includes participant incentives may reduce drop-out rates. Two types of programs for which program designers need not worry about participation are programs offering desirable goods or services to clients and mandatory

programs. Sometimes, interventions are so obviously attractive that few clients can refuse. Programs providing benefits such as housing subsidies, food, medical care, or child care usually have no need to motivate clients. Whatever the target population is for such a program, its members likely will participate once informed that the program exists. Target populations for mandatory programs are also likely to participate because their participation is required by law. Mandatory programs include substance abuse counseling for DUI offenders, traffic school for traffic law violators, and anger management classes for spouse abusers. Nonparticipation in such programs leads to financial penalties and even incarceration, so no-shows are few.

But what of the free presentations on healthy eating and the drug use awareness program for teens that go virtually unattended? Intervention programs offer to do something good for a target population, but what is “good” is not always alluring. Incentives can help. When stakeholders and evaluators suspect that a treatment or intervention in and of itself will not secure the target group’s participation, they should discuss using incentives. Fast-food restaurant coupons have been used as an incentive to attend crime prevention programs, and drug prevention programs have offered tutoring and recreation along with their curricula. Every program needs to ensure that it provides incentive or motivation enough to engage the target group in the prescribed activities.

Programs need to remove barriers to participation whenever possible. Even a willing target population, reached and motivated, may be unable to participate fully in a program given certain barriers. Such potential barriers should be weighed by the working group as it prepares the action plan. Barriers can be those of language, culture, stigmatization, or logistics. If service providers do not speak and write the *language* of the target population, potential clients may be alienated. The *cultural competence* of program staff can also be an issue. Any target population has a racial and ethnic background, and sensitivity to that culture on the part of service providers can encourage retention of clients in the program. *Stigmatization* of a program’s services raises another potential barrier to receiving those services. Participation in mental health services, for example, is viewed by many target groups as a sign of personal weakness or vulnerability. If this stigma is not overcome, target group members may decline mental health services despite implementers’ best efforts. When stigmatization of services may be an issue, the action plan might suggest strategies for alleviating clients’ fears; it might even outline some effort to change perceptions about the problem in question. An action plan may also need to address *logistical barriers*, such as lack of transportation or child care. Job training or drug treatment sites, for example, are probably not going to be conveniently located for all clients. Transportation needs are a particularly large factor in interventions with an

indigent population. Although programs cannot, of course, buy cars for clients, where public transportation is available, they can provide money for fares as one incentive to participate. If public transportation is not a good option, a program can provide a vehicle to pick up and drop off its clients. Action planners might even choose to “decentralize,” bringing services to the clients wherever they are. Like transportation, child care may need to be provided to indigent individuals so they remain motivated to participate. The working group should discuss whether any of these barriers is likely to affect its target population.

Beyond barriers of a logistical or even cultural nature is the matter of indigent clients lacking basic needs such as shelter and food and perhaps also suffering from physical or mental illness. Not infrequently, a client meeting all eligibility requirements may be in no condition to receive a planned intervention. Program implementers find themselves providing referrals or even case management services to help such potential clients before (and occasionally during) the intervention. Larger programs sometimes dedicate a unit to provide these auxiliary services. For example, when a potential client of a sobriety program remains under the influence of alcohol, delivering alcohol abuse treatment is problematic. The program that possesses its own detox unit may be successful in sobering up the individual, permitting participation in the program. The initial assessment of potential clients should make clear whether they need to become sober and also whether they have a place to live, food to eat, relative good health, and adequate education. Although it is not the job of an intervention program to provide social services, an action plan frequently includes guidelines for linking eligible clients to relevant social service agencies. The person who is without a home and unsure whether the next meal will be forthcoming is not likely to succeed in any intervention or treatment program. Thus, to be truly effective, programs whose target groups are susceptible to such material deficiencies must strive to connect these target populations with appropriate help.

APPLICATION OF THE CONCEPTUALIZATION FACILITATION STRATEGY

Example 1: A Garbage Reduction Program

In Taiwan, residents had been accustomed for four decades to placing garbage bags in designated pickup areas each and every day. Daily pickup resulted in huge amounts of garbage, to the detriment of the environment. A demonstration program proposed by the Neihou Sanitation Department was intended to determine whether a new policy—no garbage collection on Tuesdays—would

reduce the amount of household garbage set out for pickup (Chen, Wang, & Lin, 1997). Program stakeholders engaged program evaluators to tell them, with holistic thoroughness, how effective such a policy would be. For the kind of outcome evaluation these stakeholders sought, an understanding of the program theory underlying the program was necessary. Because the program was newly begun, the evaluators had an opportunity to use the development facilitation strategy to assist in clarifying and developing the stakeholders' program scope and action plan. To begin, the evaluators interviewed the program designers and other decision makers to learn about the program theory as understood by those individuals. Their basic notion of that theory, the evaluators found, hinged on the degree of unpleasantness they expected residents to experience when required to keep garbage in their homes for even one day (Taiwanese homes typically lack garbage disposals). Sufficient inconvenience and disgust, they believed, would raise awareness of how much garbage was being created, resulting in new in-home efforts to reduce its volume. The program scope and action plan that the stakeholders put together are illustrated in Figure 5.2.

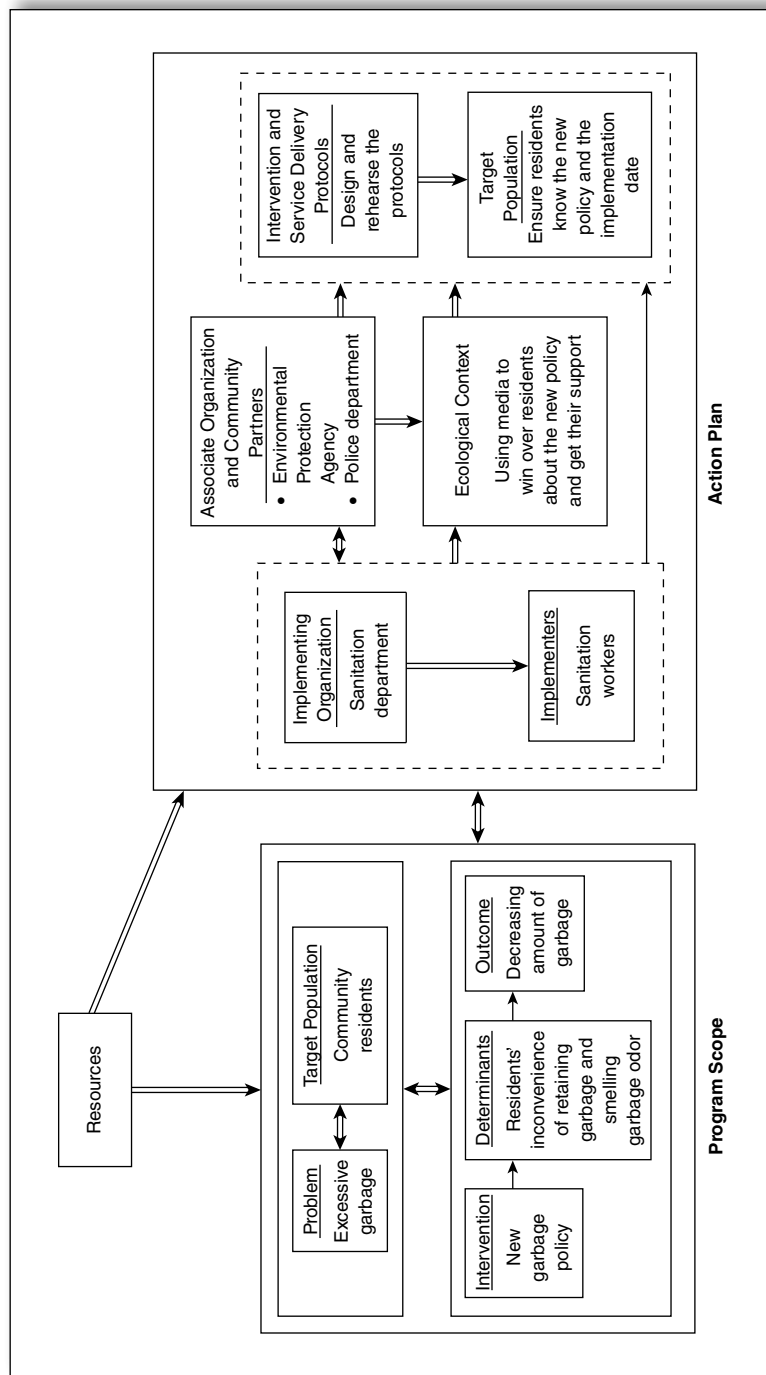
Program Scope

1. *Problem*—rapid increase in amount of garbage to be collected; ensuing environmental damage
2. *Target population*—residents of the Neihou community
3. *Outcome*—decrease in the amount of garbage to be collected
4. *Determinant*—residents' experiencing the inconvenience and odor of retaining garbage in the home overnight
5. *Intervention*—cessation of Tuesday collection of household garbage

Action Plan

1. *Implementing organization*. Neihou's sanitation department. The department had a reputation for efficiency, and it had sufficient personnel and funding to implement the program.
2. *Program implementers*. The department's regular sanitation workers, provided with training in how to implement the program.
3. *Associate organizations and community partners*. Taiwan's national environmental protection agency, the local health department, and the local police department.

Figure 5.2 Program Scope and Action Plan of a Garbage Reduction Program



SOURCE: Adapted from Chen, Wang, & Lin (1997).

4. *Ecological context.* A media-based strategy was used to address the program's ecological context. To create awareness of and support for the new policy, stakeholders planned to advertise it in local newspapers and have the policy featured in a television interview.
5. *Intervention and service delivery protocols.* Prohibition by law of disposal of household garbage on Tuesdays, to be enforced with fines. To prevent dumping at garbage collection points on Tuesdays, sanitation department employees, who were empowered to warn and then cite violators, would patrol these sites on that day. Violators would be fined for noncompliance. Pilot testing of the protocol was to take place before implementation.
6. *Target population.* Residents of the Neihou community. All affected residents were to be notified of the new policy via two waves of informational letters sent to them by the sanitation department. In addition, huge banners reminding residents of the policy were to be displayed on all major roads in the area, both before and during its implementation.

Advantages of Development Facilitation as Illustrated in This Example

The stakeholders said that the development facilitation approach helped them with the conceptualization of their program. They felt that the approach systematically organized their thoughts about the program. Stakeholders used the program scope and action plan to communicate the program to audiences inside and outside the community. The action plan was particularly helpful to them in ensuring the quality of implementation. The program scope and action plan were used as a foundation for evaluators to design and conduct theory-driven process and outcome evaluations (Chen et al., 1997).

Example 2: An HIV-Prevention Program

A funding agency concerned about the spread of HIV launched an action-planning process by offering grants to community-based organizations (CBOs) willing to undertake new HIV-prevention efforts in the African-American community. This agency brought in program evaluators, who suggested using the conceptualization facilitation strategy to help the agency's stakeholders conceptualize their grant program (Chen, 2003). The funding agency required that each participating CBO incorporate at least the following four program activities in its action plan:

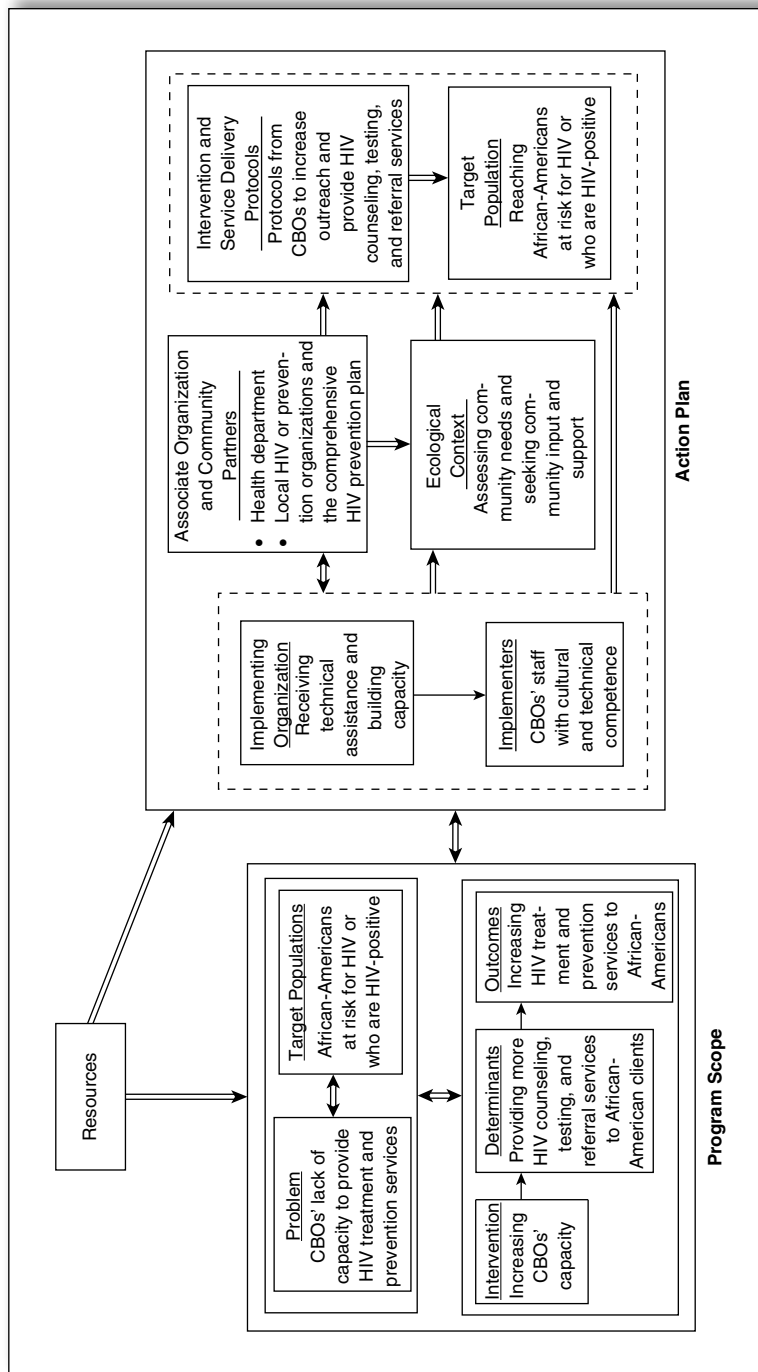
1. Through means consistent with applicable existing state and local HIV-prevention plans, provide high-risk individuals with (or help them gain access to) HIV counseling and testing, as well as with appropriate referrals for other needed services. A proposed program not strictly consistent with more comprehensive prevention plans must adequately justify its intention to address other priorities.
2. Conduct primary HIV-prevention activities such as health education and risk reduction interventions for people at high risk of becoming infected with HIV or transmitting the virus to others.
3. Help HIV-positive people gain access to appropriate early medical treatments, substance abuse services, STD screening and treatment, perinatal health care, partner counseling and referrals, psychosocial support services, mental health services, tuberculosis prevention/treatment, and other supportive services as needed. Refer high-risk clients who do not test positive for HIV to receive meaningful health education and participate in risk reduction programs and/or to other appropriate prevention services.
4. Frame all program activities and intervention messages with sufficient cultural competence and linguistic and developmental appropriateness.

The evaluators were asked to consider these requirements and facilitate a working group's planning of the grant program implementation at multiple sites. The working group included a manager and three staffers charged with bringing the grant program to fruition, plus the evaluators. The group's task was to develop a coherent action plan from the brief terms of the grant announcement. To facilitate discussion, before the working group met, the evaluators compiled a rough draft of an action plan based on the four requirements and preliminary, informal talks with the stakeholders. At its first meeting, the working group developed a program scope and action plan, summarized below and represented in Figure 5.3.

Program Scope

1. *Problem.* The upward trend in HIV transmission among minority groups
2. *Target population.* African-Americans at high risk of HIV exposure or already HIV positive
3. *Outcome.* Slowdown in the rate of HIV transmission in the target group, which has been especially susceptible to the HIV epidemic

Figure 5.3 Program Scope and Action Plan of an HIV-Prevention Program



SOURCE: Adapted from Chen (2003).

4. *Determinant.* Better access to HIV-prevention and treatment services for African-American clients who lack avenues to existing HIV testing and counseling outlets
5. *Intervention.* Provide money, technical assistance, and capacity-building services to CBOs working to reach African-Americans and guide them to HIV-prevention and treatment services

Action Plan

1. *Implementing organization.* A CBO willing to accept capacity-building and technical assistance at the discretion of, and paid for by, the grant agency
2. *Program implementers.* People demonstrating relevant technical and cultural competence
3. *Associate organizations and community partners.* Entities involved in state and local health planning affecting the CBO, plus other relevant bodies (national, regional, state, and local) supportive of HIV-prevention/treatment efforts and capable of facilitating the CBO's client referrals or curtailing duplication of efforts
4. *Ecological context.* The CBO needs to conduct needs assessment and seek community inputs and support of its services.
5. *Intervention and service delivery protocols.* To provide CBOs with protocols in counseling and testing and to provide referrals
6. *Target population.* HIV-positive and high-risk African-Americans, prioritized according to local trends (determined by needs assessment, epidemiological profiles, and state/local anti-HIV plans): intravenous drug users, men who have sex with men, sex workers, the homeless, and so forth, with *each* risk group to be reached via a tailored, concrete, CBO-designed targeting strategy

The working group decided to place the strongest emphasis on those portions of the action plan concerned with reaching the populations within the target group and linking its members to needed services.

The working group briefly discussed the type of outcome evaluation the CBOs should conduct following implementation. It reached a consensus, maintaining that such an evaluative task was beyond the CBOs' current capacities. Instead, CBOs would simply be required to collect certain evaluation data for their own use and to forward these data to the funding agency, primarily to

meet accountability needs. The working group settled on the following list of probing questions:

- Does the program reach African-Americans in high-risk populations?
- Does the program reach the target population?
- How many target population members have been served by the program?
- What kind of services have they been provided?
- What has been the quality of these services?
- Have clients referred for treatment actually received treatment?
- How many clients express satisfaction with the program?

Advantages of Conceptualization Facilitation as Illustrated in This Example

Conceptualization facilitation as deployed in the preceding sample of action planning presented five main advantages to stakeholders. First, it produced for them a coherent, systematic program scope and coherent, systematic action plan that were consistent with each other. It fostered stakeholders' understanding of, and ability to debate, the essence of the announced requirements of the grant; this helped ensure that the requirements would be reflected in the action plan.

Second, it revealed assumptions implicit in the program scope that needed to be scrutinized (and, where necessary, turned this scrutiny into action to strengthen the action plan). Looking over the program scope and plan (Figure 5.3) prompted the working group to discuss service routes. The discussion focused on CBOs and their prevention workers, who had been overemphasizing prevention services activities with clients and not emphasizing HIV counseling and testing enough. Only when the high-risk client's HIV status becomes known, through testing, can the appropriate services be determined. Strategies to increase HIV counseling and testing services by CBOs were discussed.

A third advantage of the facilitation process was its usefulness in linking the new CBO-managed activities to related activities alluded to in other funding announcements. For example, separate announcements were issued concerning capacity building and coalition development for CBOs. The model (Figure 5.3) indicates graphically a clear need to integrate the various announcements.

A fourth advantage of the facilitation process was that it helped program staff and evaluation staff to firmly understand the kind of evaluative information needed at given program stages. This understanding was a foundation from which to develop evaluation indicators—and an excellent evaluation design.

The fifth advantage of the conceptualization facilitation strategy used here was a by-product of the process: insights as to promising foci for future programs. For example, although the grant announcement cited in this example was silent on the issues of removing barriers, motivating clients, and building contextual support for intervention, the model of conceptualization facilitation shows that these are, indeed, important domains deserving incorporation within action plans. Thus, the working group's discussion of future action planning was a conversation that has likely informed each member's work with action plans.

THE PILOT-TESTING APPROACH

The pilot-testing approach falls under the troubleshooting strategy in the typology of program evaluation means and ends (see Table 2.1). Pilot testing determines the field feasibility of an action plan and thus is a way to avert the implementation of a significantly flawed plan. Often, once an action plan is completed, the program director and implementers feel pressure to implement the plan immediately. Pilot testing is valuable when these stakeholders want quick feedback to use in fine-tuning their plan before implementation. Because timeliness is likely to be paramount, a pilot test must be flexible as to research methods; this flexibility even extends to testing only parts of a program rather than the entire program. More evidence of pilot testing's flexibility is its use of samples much smaller than those demanded by traditional evaluation. Strictly speaking, the rapid inquiry of the pilot test is more a development tool than a formal assessment tool.

Defining Pilot Testing

The term *pilot testing* is often associated with tests of the reliability and validity of questionnaires or measurements. In this book, however, the *pilot testing* does not refer to measurement alone. Rather, it refers to a feasibility study, a small-scale field trial of an action plan conducted rapidly—over a few weeks or months—to assess and improve the implementation of the full-scale intervention. Key to the success of pilot testing are the evaluator's timely analysis of data and presentation of findings to stakeholders. A pilot test can include testing of instruments, but its main purpose is to discover problems that might arise during implementation of an action plan. The characteristically small sample size involved in pilot testing can shrink to just a handful of subjects when the program under review focuses on hard-to-reach clients such as homeless people and drug addicts.

Furthermore, it is important to distinguish pilot testing from related concepts like *pilot study* or *demonstration project*. These activities involve far more elements than does pilot testing. Pilot studies and demonstration projects characteristically test all aspects of a program and evaluate how the various stages of that program work together, thereby testing the program's effectiveness. Their ultimate concern with program effectiveness encourages the use of rigorous methods and, in turn, the diversion of significant resources.

Conducting Pilot Testing

Four principles are especially significant to the program evaluator preparing a pilot test to generate information that can improve action plans:

1. *Pilot testing requires actual implementers and clients to participate in trials.* Feedback from the very implementers involved deeply in the day-to-day activities of the program offers firsthand information about looming problems with implementation, as well as educated guesses about managing problems. Feedback from the clients whose very lives can be changed by a program produces genuine insight into what makes, for them, a satisfying intervention delivery. In earlier chapters, it was emphasized that action plan development is the realm of experts and top officials of organizations; however, implementers and clients may have their own, very different and very valuable, perspectives on a program.

2. *Pilot testing relies on small but nevertheless typical samples.* To provide feedback promptly, pilot testing usually relies on a small group of clients and implementers. For best results, these participants should be typical of their groups. For example, if a majority of a job program's clients are expected to be drawn from the persistently unemployed, then the handful of clients in the pilot test should also be persistently unemployed. If instead these pilot-test subjects were relatively experienced in keeping a job, the pilot-test results would be highly misleading. A good rule of thumb for pilot testing is to avoid extremes, such as overly enthusiastic or apathetic, or overachieving or underachieving, target group members. Another rule of thumb is that, when a sample implementing organization is needed for the pilot test, one should select an organization typical of the ones that will conduct the full implementation. If mom-and-pop CBOs will implement the program, do the pilot testing with a mom-and-pop organization, because results obtained for a large, complex, sophisticated organization would be difficult to generalize to the complete program in its planned form. The same principle should be applied when pilot testing a delivery mode, ecological context, or other component.

3. *Methods of gathering data in a pilot test must be flexible.* With a small sample and short time frame, the pilot test is, in most instances, better off collecting its data with methods like the focus group, interview, survey, site visit, and so on. There is no need to apply rigorous methods like the randomized experiments often used in a pilot study or demonstration project.

4. *Pilot-test findings should be used only for program development purposes.* Pilot testing is of a developmental nature. Because it tests only parts of a program, results should never be used as evidence of a program's effectiveness. The pilot test asks whether a program shows signs of major implementation problems, not whether it will be effective in the end. Interpreting the results of pilot testing is made simpler by this rule of thumb: When a pilot test suggests an action plan is substantially flawed, full-scale implementation of the plan likely will be marred unless the flaws are resolved in the action plan beforehand. On the other hand, when pilot testing suggests that the action plan works well, the plan may—or may not—generate a successful full-scale implementation. Positive results from pilot tests simply are not indicators or evidence that an implementation will be of high quality or that an intervention will be effective. Such evidence comes from process evaluation and outcome evaluation, not pilot testing.

Designing Pilot Testing

The key to designing pilot tests is to mimic exactly the program activities and processes planned for the full-scale implementation. For example, for interventions that, when implemented, will comprise several sessions over a period of time, the pilot test ideally would involve an identical schedule: same session length, same day of the week, same time span, and same setting. The action model framework can be consulted as a systematic means of considering each component of the action plan to be tested. It is up to the stakeholders to determine just what information will be collected via a pilot test. The evaluator may present them with certain guidelines, however, for selecting or passing over topics.

In general, the following questions need particular attention during pilot testing:

- Can the intervention be implemented in the field as intended?
- Can implementers anticipate encountering certain problems delivering the intervention?
- Will clients be receptive to the intervention or resist it?
- Do any of the program's organizational procedures impede the implementation process?

Intervention and Service Delivery Protocols. Protocols are the components most frequently evaluated by pilot testing. They are tested by delivering an intervention to clients in the planned setting. Although comments from implementers and clients are something of a reality check, evaluators can also choose to observe in person the delivery of a service, watching for potential problems. When clients and implementers are queried in the course of pilot testing, the following issues should be brought out in interviews or surveys:

- Is the dosage of the intervention satisfactory?
- Does the intervention require too much time? Conversely, is it too brief to be effective?
- How much difficulty do implementers experience in delivering the service?
- Do clients find it difficult to follow the language or procedures of the intervention?
- Does the intervention setting help or hinder service delivery?
- Is the intervention schedule acceptable to clients and implementers alike?
To illustrate, consider a day care center for working mothers, sponsored by a welfare program that has chosen to close the center at 5:30 PM each day. Because many of the women use public transportation and deal with its vagaries, this schedule is not acceptable to them.

Deciding how to implement an intervention in the field is one of the most challenging tasks facing a new program. Prospective clients and implementers alike can contribute to the decision by giving feedback about the practicality of the planned mode of service delivery for the planned intervention in the planned setting. For example, one action plan called for hiring a professional outreach worker to bring a prevention message to a high-risk neighborhood. But when speaking with a handful of prospective clients during a pilot test, the program evaluator heard their opinion that an outsider would be given little opportunity to interact with the neighborhood's residents. This finding prompted revision of the action plan, which eventually stipulated the involvement of outreach volunteers from the neighborhood.

Target Populations. The reaching and recruiting of members of a target population is another important focus for pilot testing. No program will survive for long without clients, and a pilot test is an opportunity to ensure that clients can be obtained via a proposed recruitment strategy. Take as an example a child abuse prevention intervention targeted at parents living in inner-city high-rises. Its proposed recruitment strategy is complicated: The cooperation of building managers is required, and doors must be knocked on one by one to introduce households to the program and extend individual invitations to participate.

Stakeholders who decide to pilot test the recruitment strategy can learn whether building managers support the program, whether parents are home during the day, whether the population of young children in the buildings is significant, and whether parents feel open to the notion of attending the program or would feel stigmatized by it. If pilot testing shows the strategy is likely to fail, stakeholders have a chance to devise a stronger strategy. This alone could be the difference between a foundering program and a thriving one upon full-scale implementation.

When action planners see a need to use incentives to recruit clients, a pilot test can show whether the selected incentives will actually motivate individuals to complete the program. Is a storybook incentive enough to bring children to dental health sessions? Pilot testing should be able to answer the question quickly. (Sometimes the results can be unexpected.) Any barriers to program participation can also be scoped out through pilot testing. A mental health program for the poor, for example, might discover through pilot testing that eligible people believe that such services carry a stigma, discouraging them from participating. The action plan could be modified to address this information.

Implementing Organization and Implementers. Pilot testing an implementing organization is more difficult than testing other action model components. In many cases, evaluators resort to site visits instead. Such visits are an opportunity to gather data about organizational capabilities, indirectly approaching the issue. General inquiries during a site visit could include the following:

- Does the implementing organization have the skills, resources, and commitment necessary to implement the intervention?
- Are the implementers qualified to deliver services?
- Can the implementing organization build and sustain collaborations or linkages with related organizations to facilitate delivery of the intervention?

Ecological Context. Sometimes client participation hinges on the ecological context of a program, notably whether family and friends support an individual's enrollment in that program. Stakeholders must ask whether such support can be counted on, and a pilot test can often give an answer. For instance, in a plan for a delinquency intervention program for juveniles, if it is acknowledged that parents' involvement and support are key to the success of the program, pilot testing this ecological context constitutes a valuable safeguard against the absence of these ecological elements. The pilot test will indicate whether parents of the targeted juveniles are supportive. After analyzing and compiling the data generated by observation and interviews, the evaluator works with stakeholders to revise and finalize the action plan prior to implementation.

A good example of the course of pilot testing is found in a school-based cancer prevention program, “Key to Health,” which proposed using five weekly, 90-minute sessions in which teachers taught adolescents about the role of a low-fat, high-fiber diet in preventing some cancers and supported their efforts to follow such a diet (Wallin, Bremberg, Haglund, & Holm, 1993). A pilot test saw two portions of the Key to Health curriculum offered to students during a 3-month period. Essentially, these test sessions gave the students an opportunity to become aware of and reflect on their own eating habits and perhaps experience a healthful, self-initiated change in diet. Each student set an individual dietary goal and made changes based upon it. At the end of the third month, information about the program’s operation was collected from those teachers and students who had participated in the test sessions. Of the students, 49 completed self-administered questionnaires and 8 students joined a focus group moderated by a program evaluator; the teachers completed both structured interviews and a survey. Results suggested that the tested Key to Health program had worked well and could be integrated readily into existing curricula (into a home economics course, perhaps). But the pilot testing also indicated a discrepancy between students’ and teachers’ views of self-efficacy as a force behind healthful dietary changes. Key to Health stakeholders decided that any setbacks likely to stem from this discrepancy could be overcome with teacher training.

THE COMMENTARY OR ADVISORY APPROACH

Evaluators are not always asked to provide formative research or development facilitation. Instead, stakeholders may ask them (or other experts) simply to comment on a completed program scope and action plan. The stakeholders are especially likely to seek advice on how these might be improved. In a situation like this, program evaluation occupies the role of troubleshooter; indeed, the commentary approach falls under the troubleshooting strategy in the practical typology (see Table 2.1). The conceptual framework of program theory, including the change model and action model alike, is a beneficial tool for evaluators when it comes to commenting and advising. It prompts a number of questions whose answers are clues to the quality of the program scope and action plan. Evaluators might ask this set of questions to locate those points on which they should comment.

Questions to Inform the Evaluator’s Commentary on a Program Scope

- Does the program scope specify target populations and systematically explain the intervention, determinants, and goals/outcomes?

- Are all elements present in the program scope adequately specific and apparently justifiable?
- Do the relationships that the program scope assumes to exist among these elements stand up to scrutiny?
- What procedures have been used to ensure the various stakeholders' understanding and support of this program scope?

In fact, the first question an evaluator might ask when critiquing an action plan is, Does a program scope exist to guide the development of the action plan? If not, the evaluator should steer the stakeholders to adopt a program scope, which they can then draw upon to revise their action plan. The following recommended *further* questions are generated from the action model conceptual framework.

Questions to Inform the Evaluator's Commentary on an Action Plan

About the Implementing Organization

- Does the implementing organization have the experience and the capacity to implement the intervention?
- Is the implementing organization experienced in working with the target group, and is it sensitive to this group's culture and needs?
- Has the implementing organization earmarked the necessary resources and personnel to implement the intervention?

About Intervention and Implementation Protocols

- Does the action plan include an intervention protocol specifying which curricula and activities the intervention will comprise?
- Does the action plan include an implementation protocol specifying the setting for service delivery and the procedures to use in delivering services to clients?

About Implementers

- Have reasonable minimum qualifications for implementers been clearly specified?
- Will training be provided to implementers?
- Have procedures (e.g., training) been planned to ensure implementers' cultural competence?

About Associate Organizations/Community Partners

- Has the implementing organization identified pertinent organizations with which it might profitably collaborate on service delivery?
- Has the implementing organization developed strategies for working with associate organizations to facilitate service delivery?

About Ecological Context

- Will support from family, friends, and/or co-workers be required for clients to succeed in an intervention? If such support is required, what strategies and procedures are in place to secure it?
- Will community support of the intervention program be required for the program to work? If it will, are strategies and procedures in place to secure such support?

About Target Populations

- Are the eligibility criteria for potential clients clearly defined and practical to implement?
- Have precise and feasible strategies been developed to reach the target populations?
- Are sufficient incentives in place to persuade clients to participate in the program?
- Does the program adequately understand and address the barriers that may come between target population members and program participation?
- Can the program recruit, with reasonable effort, a sufficient number of clients?

SUMMARY

The four evaluation approaches covered in this chapter—formative research, conceptualization facilitation, pilot testing, and commentary or advisory—are means by which an evaluator can assist stakeholders as they develop action plans.

Formative research generates background information that stakeholders can use in creating an action plan. Conceptualization facilitation allows the evaluator to facilitate the stakeholders in clarifying or developing an action plan.

With an action plan already in hand, the approaches most useful to typical stakeholders are pilot testing and commentary or advisory, both of which work

to fine-tune an action plan. The commentary approach is a low-cost alternative but has a disadvantage in that it does not use hard data from the field. Even experts' comments and suggestions can be too general and subjective. Although useful in its own way, the commentary approach does not work as a substitute for pilot testing when pilot testing is needed. The most important advice this book can give along these lines is to use pilot testing—if *at all* possible—to fine-tune any start-up program.

Based upon the program scope and action plan, stakeholders can figure out the manpower and budget needed for implementing the program. The full version of the program plan is very useful for program management and evaluation purposes.

Making program evaluation part of a program from its earliest planning stage is of major benefit. Evaluators bring broad expertise to the table; their know-how extends even to stakeholders' distant evaluation needs, those that will arise in the final stages of program development and after. Process evaluation and outcome evaluation are the options most likely to meet these late-stage needs, so in some circumstances, it may make sense to include in the action plan a schedule of future processes and/or outcome evaluation (perhaps along with a statement of stakeholders' expectations for such evaluation). The details of constructive, conclusive, and hybrid process evaluations are discussed in Chapters 6 and 7 and those of outcome evaluation design in Chapters 9 to 12 of the book.

QUESTIONS FOR REFLECTION

1. What is an action plan? How does it relate to the action model framework discussed in Chapter 3?
2. How is formative research useful to developing the action plan?
3. What might happen if program planners do not consider the three main factors—technical expertise, cultural competence, and manpower and other resources—when planning a program? Can you think of any experiences you may have had in which an implementer or practitioner was not culturally competent?
4. Why is it important to consider the ecological context when planning a program? How does microlevel support differ from macrolevel support?
5. How does the intervention protocol differ from the service utility protocol?

6. List real-life examples of each mode of delivery. Discuss which mode of delivery you prefer.
7. What role does cultural competency play when identifying a target population?
8. Using an actual intervention program, identify the target population, goal and outcomes, determinant, and intervention.
9. Why do you think African-Americans were the target population of the HIV-prevention program discussed in “Example 2”? Do you think it would have been better to further define the target population (e.g., African-American males)? Why or why not?
10. What is the main purpose of pilot testing? What information can be gathered by pilot testing? Did the text’s definition differ from what you originally thought *pilot testing* meant? Is it always necessary to pilot a program before implementation? Explain.
11. Compare and contrast pilot testing discussed in this chapter with relevancy testing, discussed in Chapter 4 on program scope.

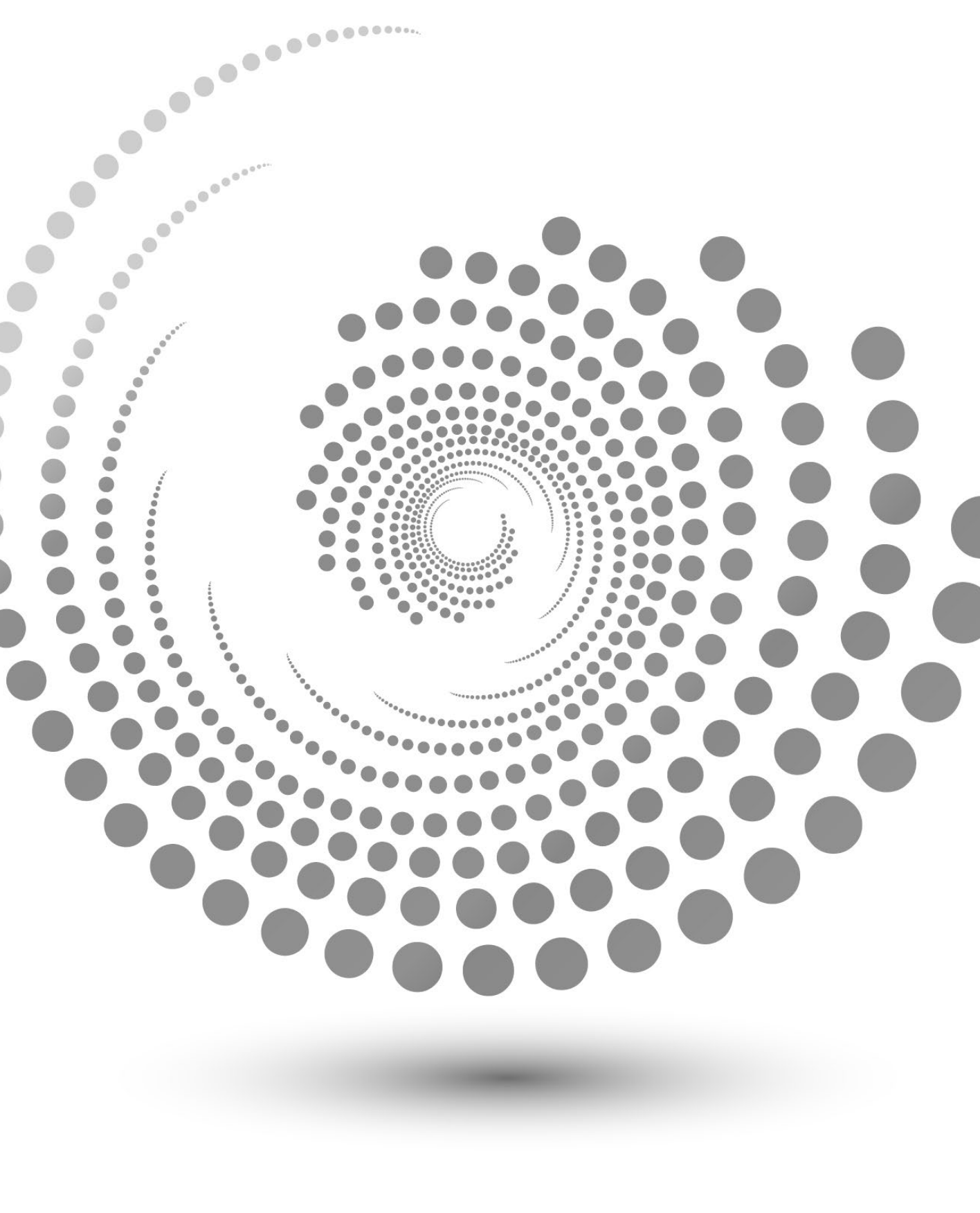
PART III

Evaluating Implementation

Implementation is where the action is in programs. An implementation must be a success before a program can be considered effective; without appropriate implementation of the program plan, there can be little expectation that the program will succeed. Implementing a program plan is usually complicated and challenging. Program evaluation is a good tool for stakeholders who want to ensure that the implementation of their program is being carried out as intended. Evaluation of implementation processes has been called implementation evaluation or process evaluation. In discussing implementation evaluation, it is important to note that implementation has two stages: the initial implementation and the mature implementation. Distinguishing the stages from one another is vital because stakeholders' evaluation needs are distinct at each stage. What stakeholders are seeking at the initial implementation stage is quick feedback for use in developing their program; thus, the evaluation information obtained is largely for internal purposes. In evaluating the initial implementation, strategies, approaches, and research methods need to be flexible and expeditious. Constructive evaluations are useful in this stage.

In contrast, in the mature implementation stage, evaluative information is needed to meet accountability requirements (as well as to shape stakeholders' discussion of long-term strategic improvement of their program). External audiences such as funding agencies expect stakeholders to provide credible evidence of the success of a program's implementation. For this reason, evaluation procedures and research methods used to assess mature implementation must emphasize rigor. Scientific rigor typically requires a relatively greater investment of time, so evaluations of the mature implementation stage usually take longer to design and conduct. Conclusive process evaluation or hybrid evaluation is particularly useful at this stage.

When stakeholders approach an evaluator about process evaluation, the parties initially need to discuss the stage of implementation in question and the purpose for the process evaluation. Communication about these two aspects, in particular, will allow the selection of an appropriate evaluation strategy and approach for the task. In Chapter 6, the strategies and approaches appropriate for evaluation of an initial implementation are presented, whereas Chapter 7 focuses on procedures best used to evaluate a mature implementation.



CHAPTER 6



Constructive Process Evaluation Tailored for the Initial Implementation

An evaluator brought in during a program's initial planning stage may have previously evaluated the program's plan. If so, his or her familiarity with the intervention is a boon for the design of the program evaluation. If not, it is well worth the investment of the evaluator's time, before designing and conducting the evaluation, to learn how stakeholders have conceptualized their program.

The initial implementation stage remains a formative one in which the program is fluid. The procedures and rules governing implementation are frequently revised. During this stage, many programmatic concerns are being addressed simultaneously: recruiting and training implementers, establishing service delivery procedures, reaching potential clients, contacting related organizations, dealing with unexpected crises, and so on. We have seen how a high-quality program plan and the use of pilot testing reduce the number of problems encountered. Even with these safeguards, however, difficulties can arise. When having the initial implementation evaluated, stakeholders are looking for quick feedback about significant setbacks with implementation in the hope of overcoming them efficiently and averting damage to the program as a whole. Thus, the type of evaluation they want is *constructive evaluation*, as discussed in Chapter 1. In addition, in the comprehensive evaluation typology (Table 2.1), constructive evaluation corresponds to two strategies that are highly recommended for evaluation of initial implementations: the troubleshooting strategy and the development partnership strategy. The *troubleshooting strategy* returns feedback from the field rapidly, reporting suspected problems with the implementation and giving insight into possible solutions. The *development partnership*

strategy is a response to stakeholders' concerns that a program and implementation may be overly fluid and vague; it involves a program evaluator functioning as a partner in development work.

As indicated in the typology (Table 2.1), the troubleshooting strategy covers two evaluation approaches: the formative evaluation and the program review/development meeting. The evaluator can use *formative evaluation* to collect firsthand information about potential implementation problems and their management. The *program review/development meeting* is more useful for building consensus among program staff as to the implementation problems encountered and the remedial actions available. In slight contrast to the troubleshooting strategy, the development partnership strategy and approach fit any program that seeks frequent input from evaluators during its planning and development phases. Below, the nature of each strategy and approach, and its pros and cons, are discussed. The discussion provides a guide to selecting and applying appropriate procedures for evaluation of initial implementation.

THE FORMATIVE EVALUATION APPROACH (UNDER THE TROUBLESHOOTING STRATEGY)

Troubleshooting is an expeditious strategy for developing stakeholders' awareness of major problems in a program or barriers to its implementation. The troubleshooting strategy allows evaluators to systematically and rapidly gather information about the existence and possible sources of problems; it also equips them to facilitate stakeholders' efforts to remedy problems. Troubleshooting can be conducted at any point following a program's implementation, but it is especially useful during the initial implementation stage.

One popular evaluation approach associated with this strategy is *formative evaluation*. Formative evaluation employs flexible research methods to assess barriers to and facilitators of implementation, enabling stakeholders to troubleshoot problems. Formative evaluation is different from the *formative research* discussed in Chapters 4 and 5: Formative research provides background information to further stakeholders' design of a program, whereas formative evaluation is a development-oriented evaluation applicable once a program is formally implemented. Because formative evaluation is conducted only with formally implemented programs, it is also different from the pilot testing described in the preceding chapter. In fact, large and complicated intervention programs usually need formative evaluation *and* pilot testing. (Generally, small and/or straightforward programs require only one or the other.) Pilot testing involves pieces of programs rather than whole ones, or it tests smaller-scale versions of the proposed program. Although pilot

testing is useful in preparing for initial implementation, formative evaluation is still needed in this stage to ensure the success of the implementation.

Timeliness and Relevancy

Good formative evaluation must meet two criteria simultaneously: timeliness and relevancy. Timeliness means that an evaluation keeps to the stakeholders' time frame. Because stakeholders are, at this point, in need of quickly collected information they can use to continue efficiently down the road toward mature implementation, an evaluation that is too slow paced will not serve. A great deal of the value of formative evaluation lies in its ability to present information to stakeholders quickly. The value of formative evaluation also resides in its ability to identify crucial implementation problems likely to affect the quality of the program overall. The action model conceptual framework shows evaluators where to focus in their search for such problems.

Research Methods

To ensure its relevancy and timeliness, formative evaluation tends to employ research methods that are flexible and can be tailored to particular evaluation circumstances. Research methods often used in formative evaluation are focus groups, participant observation, key informant interviews, and small-scale surveys. These suit formative evaluation's focus on programmatic inquiries: Can implementers reach the intended clients? Are implementers having difficulty delivering services? Are clients receptive to the intervention? Does the community support or oppose the implementation? The answers to these questions direct stakeholders in fine-tuning and managing the program for success down the line. Program directors and implementers frequently adopt the results of formative evaluation to revise the implementation process during the initial implementation stage when program structures are not yet firmly established and modifications are relatively easy to make. Formative evaluation provides timely information to serve programmatic needs.

A hypothetical program demonstrates the usefulness of timely, relevant formative evaluation. The clients of a fairly new family counseling program for Asian immigrants are frustrated with what they perceive to be the counselors' lack of understanding of Asian cultural beliefs and family values. Evaluators can document this problem quickly through a formative evaluation that uses either focus group meetings or interviews with the immigrant clients. Once detected, the problem can also perhaps be solved quickly. Trying to use time-consuming,

large-scale, rigorous research methods with a large representative sample could result in an evaluation that finally wraps up *after* numerous clients have become so frustrated that they opt out of the program. There is a place in process evaluation for rigorous design and research, of course, but it is not in the initial implementation stage. (Chapter 7 will explore the comparative rigor of process evaluation during the mature implementation stage.)

Steps in Applying Formative Evaluation

The formative evaluation approach is applied in six basic steps. They are presented here, along with the general principles of formative evaluation pertaining to each one.

1. Review Program Documents and Note Underlying Assumptions

To be sensitive to issues involved in the development of a particular program, evaluators must know the program, and its purpose, in detail. Evaluators who participated in developing a program rationale and program plan have an advantage when it comes to carrying out formative evaluation of that program: They are aware of the assumptions that underlie the stakeholders' decisions. However, the evaluator invited to carry out formative evaluation after the closing of the planning process can acquire the necessary information from documents and interviews with stakeholders. Without mutual understanding of the program between evaluator and stakeholders, the quality of the formative evaluation is in jeopardy.

2. Identify the Program Elements Crucial to Successful Implementation and Determine Which May Be Vulnerable

The need for timeliness makes it impossible, during process evaluation, to examine every aspect of a program. Evaluators need to work with stakeholders to determine which parts of the program are likely to be most vulnerable and thus deserve additional attention. Using the action model conceptual framework, the evaluator can facilitate brainstorming by stakeholders about vulnerabilities that may call for an intensive check.

3. Select Well-Suited Data Collection Methods

Again, formative evaluation uses research methods that are flexible and provide quick feedback. Participant observation, key informant interviews,

focus groups, site visits, record reviews, and small-scale surveys are some popular tools. Often, an evaluator must tailor the research method somewhat to fit the evaluation circumstances. For example, when surveys are employed in formative evaluation, they may not embody ideal survey methodology—notably, the use of a large, representative sample. To work within the stakeholders’ time frame, smaller samples must do. Formative evaluation deals with programmatic issues, which are relatively robust and easily documented (unlike proof of causal links between variables, the object of traditional research). Within the field of program evaluation, information from focus groups and small survey samples is perfectly capable of raising the necessary red flags if a program is facing real problems. For example, a focus group might voice a consensus opinion that waiting times for services are too long and frustrating, and a number of members may add that implementers’ rudeness is fueling resentment. Whatever methodological limitations “taint” the information, it still points to a problem needing immediate attention.

4. Identify Problems

In the information the evaluator obtains, any elements or activities of the implementation suffering difficulties should show themselves. The findings can quickly be related to the stakeholders.

5. Probe for Sources of Problems to Help Stakeholders Choose Remedial Action

Formative evaluation becomes more useful to stakeholders when it goes on to provide information about a problem’s source and strategies that might resolve the problem. This is why probing for the reasons behind a problem is an important part of formative evaluation. For instance, the evaluator who finds that clients are unreceptive to implementers will immediately ask why they are unreceptive. Do they feel ill at ease with implementers because the program’s staff members seem inadequately trained, or insensitive to culture and language, or overly hurried? Answers to the evaluator’s probing questions generate timely feedback, providing a base for remedial action by stakeholders.

6. Submit Findings to Stakeholders and Document Changes They Make Based on the Findings

The results of formative evaluation will likely prompt stakeholders to modify the program; ideally, formative evaluation does what is needed to ensure the

program is implemented appropriately. Unfortunately, the emphasis on taking action to fix problems contributes to a tendency among program staff to be less than diligent about documenting modifications to the program plan. Eventually, the written program plan and other documents no longer reflect the reality of stakeholders' intentions. Such neglect can create many problems when a program is later assessed on the merits of its implementation or on its effectiveness. It is highly desirable for evaluators to work with stakeholders to immediately and systematically document all important changes and revise the program plan accordingly.

Four Types of Formative Evaluation

Formative evaluation frequently takes one of four forms: on-site observation and checking, focus group meeting, intensive interviews, and systematic scanning. Below, the nature of each of these forms is discussed, and the steps in using them are listed.

1. On-Site Observation and Checking

In on-site observation and checking, the evaluators themselves participate in a program, or else observe the implementation process, to identify major implementation problems (if any) and probe their causes. This kind of observation entails witnessing the service delivery processes and then interviewing clients and implementers. Program directors and staff are quickly informed of findings to facilitate their decision making.

On-site observation and checking cannot be done without some preparation. The evaluator does not simply jump into the field with no conceptual grasp of the program's intentions and limits just to observe whatever transpires. To get useful information to stakeholders, evaluators need to be familiar with stakeholders' ideas about the program plan. The action model conceptual framework can help the evaluator systematically review these ideas with the stakeholders. In the course of discussion, the evaluators should probe for areas of the program stakeholders might consider to be weak and subject to potential implementation problems. This task is important because it ensures that, at a minimum, the evaluation will provide the information of most interest to stakeholders. However, the evaluator should not feel confined to only those potential weaknesses remarked on by stakeholders; once in the field, the evaluator must feel free to investigate other issues emerging from observation or interview.

During on-site observation and checking, there is no need to check for all potential problems before communicating with stakeholders. Time is of the essence, so when an implementation problem (and perhaps its likely source) has been identified, the evaluator passes the information along immediately to the program director and implementers. To withhold the information would be to deny stakeholders time they could have used to develop a resolution. Nor is there any requirement that on-site observation and checking be a one-shot evaluation. The approach can be applied for as long as needed; in general, as compared to smaller programs, larger programs take longer to reach a state of mature implementation, meaning that on-site observation and checking of a larger program will reasonably cover a longer period.

An Example of On-Site Observation and Checking. Shapiro, Secor, and Butchart (1983) studied a leadership- and management-training program designed for women working in higher education; their effort provides a good example of useful on-site observation and checking employed to strengthen a program in its initial implementation stage. The training program's purpose was to develop positive self-concepts in the women that might help move them into administrative positions. Most participants were support staff, entry- or midlevel administrators, and faculty. The studied interventions included brown-bag seminars, a case study workshop, and a leadership and management clinic, which evaluators joined as participant-observers. Evaluators also conducted surveys of clients' satisfaction with these activities. The evaluators identified three major problems and described them for the program developer in ad hoc summaries. The problems were segregation within the target group, confusion about co-facilitators' duties, and a feeling that presentations were somewhat nebulous. At the first meeting, participants sat with and worked with other participants in the same occupation, so support staff did not mix with administrators or faculty; most viewed this as elitism. Made aware of this difficulty, program staff used mixed groups in subsequent meetings. The sense of occupational segregation and elitism was no longer an issue after this change.

Furthermore, program stakeholders had hoped that the co-facilitators they hired would evolve into strong leaders. But in interviews with the co-facilitators, evaluators found that they felt unprepared for their tasks because they believed their role was too vaguely defined. This finding led to the preparation of formal written descriptions of the co-facilitators' duties during each phase of the program. Follow-up interviews showed that making these roles explicit did much to relieve the co-facilitators' concerns and anxieties.

The program developer made three presentations during the case study workshop. Word reached the program developer of comments from the participants that a written summary of the presentations would solidify the content and make the presentations more valuable. In response, the program developer produced an agenda, provided written outlines of content, and listed related topics for group discussion. This example shows how formative evaluation can lead to effective program changes.

2. Focus Group Meeting

The focus group is another method well suited to formative evaluation (Krueger & Casey, 2015). The focus group embodies an interactive strategy for gaining knowledge of the perceptions, experiences, and beliefs of a small group of people about a topic or experience with which they are familiar. The knowledge is generated through discussions guided by a moderator; program evaluators can make good focus group moderators. Focus groups should be flexible, relatively simple to conduct, and cost-effective. Moderators should use a group's discussions to probe clients' and/or implementers' perceptions of the strengths and weaknesses of a program. Through the focus group meeting, thorough and detailed information can be acquired that a pen-and-paper survey does not elicit. That is the advantage of focus groups, but the method has a disadvantage, too: It does not yield generalizable numbers, such as exact percentages of people holding a particular belief or encountering a particular experience. But, as has been discussed, this may not be a pertinent matter, especially during the initial implementation stage. Furthermore, if necessary, focus group data collection can be augmented with a survey or other quantitative method.

An Example of a Focus Group Meeting. Quantock and Beeynon (1997) evaluated an osteoporosis awareness program using process evaluation and a focus group. The team's purpose was to see whether the program was meeting both patients' perceived needs and their medically identified needs. Sixteen female patients living within 10 miles of the hospital were asked to join the focus group. Each woman received an explanatory letter and a list of the five topics (related to the program's purpose and objectives) the focus group would discuss. Eleven patients accepted the invitation; transportation was provided as needed. Participants agreed to the focus group confidentiality policy. An independent clinical psychologist facilitated the focus group; implementers of the osteoporosis program were not present. The discussion lasted about 45 minutes, after which refreshments were served.

Although data from the focus group meeting showed the program was addressing several needs the patients had, it also showed that patients desired improvement in four areas. They were dissatisfied with the staff's general level of knowledge and with the length of time required for diagnosis. They felt confident about the hormone replacement therapy they were receiving, but they were less satisfied with regard to the benefits of other employed therapies, such as bisphosphonates, a class of drugs that prevent the loss of bone mass. Above all, the focus group participants felt great fear about the future. They attributed this, in part, to what they saw as an inappropriate focus by program staff on disabilities stemming from osteoporosis rather than on the patients' remaining capacities and their potential to retain these capacities.

Program planners implemented four changes in light of the focus group discussion: (a) The program staff's professional awareness of osteoporosis and its management was more strongly emphasized, with growth of this knowledge base becoming an ongoing requirement; (b) information provided to patients about various treatment options was made more comprehensive and equitable, and information about the benefits of treatment was given added emphasis; (c) staff members were reminded of the empowering effect of a positive attitude toward patients' future health, and realistic lifestyle advice and practical information about the risk of bone fractures were made more readily available to patients; and, finally, (d) a requirement was added to continue evaluating the program in order to address the changing needs of patients.

3. Intensive Interviews

During a formative evaluation, intensive interviews with individuals can be as helpful as focus group discussions. Face-to-face interviews of clients and/or program staff such as managers and implementers are a good means of collecting data, especially when evaluators ask probing questions to follow up on more complicated issues.

An Example of Intensive Interviews. Hawe and Stickney's (1997) evaluation of one coalition provides an example of how the intensive interview can gather data useful in formative evaluation. This food policy coalition had an ambitious purpose: to improve the food supply and to improve cooperation among organizations to facilitate the provision of adequate, nutritious food. After 12 months of operation, the coalition saw itself as floundering and sought an evaluation prior to trying to develop strategies to boost its productivity. It hoped for the kind of feedback that would be a catalyst for change, directing whatever restructuring of the program was needed. The coalition wanted the

evaluation to be finished and the feedback information in its possession within 3 months. The evaluation literature was the basis for the researchers' decision to focus the evaluation on these areas: coalition members' perceptions of the role and responsibility of the coalition, patterns of attendance at coalition meetings, the members' degree of involvement in and satisfaction with the work, the characteristic decision-making process of the coalition, members' expectations about outcome efficacy, and members' suggestions to improve productivity. There were 21 members and former members of the coalition who agreed to be interviewed. Members were sent in advance a self-administered questionnaire regarding the coalition's effectiveness in achieving its goals.

The data that were obtained revealed a few main problems: an insufficient mechanism for attracting new members, conflicts among the perceived roles of the coalition, and a notable lack of confidence in the coalition's capacity to achieve its goals. The evaluation feedback and follow-up discussion allowed the current members to implement several changes in the coalition's operations. In sum, the coalition's structure was recast, stronger mechanisms were created for realizing goals, and incentive management (ways to enhance benefits and lower costs to the diverse parties involved) was strengthened.

4. Comprehensive Scanning

A formative evaluation of a large program usually requires using more than one research method to acquire data in the field. This kind of formative evaluation is called comprehensive scanning. Comprehensive scanning rapidly identifies major implementation problems and otherwise scans for opportunities to enhance a program. Scanning differs from on-site observation and checking in two ways. The first is scale: Scanning is typically applied with large programs or programs with multiple sites operating simultaneously. The second—which grows out of the difference in scale—is the method of data collection. Larger-scale evaluations entail difficulties in using evaluators' participation or observations as a central source of data. Scanning usually relies on simultaneous deployment of several data collection methods, such as record reviews, telephone conferences, emails, site visits, interviews, and surveys. But like other formative evaluation methods, scanning depends on the stakeholders' program plan to guide evaluation activities. Principles discussed above (see "On-Site Observation and Checking") are thus applicable to scanning as well. The action program conceptual framework serves as a guide to the important focal areas during scanning, or as a discussion map for evaluator and stakeholders to use as they determine which elements of the program plan are most likely to be vulnerable and in need of close observation. Data taken from the field may

raise other concerns, which evaluators should pursue if possible. Periodic forwarding of information to stakeholders allows them to take remedial action in the timeliest fashion.

Although scanning is most often needed during formative stages of program development, it is sometimes employed to ensure that a program is operating properly in its mature implementation stage. A fully mature program may be well served by a *permanent evaluation system* that monitors the primary areas of implementation. It is important to understand that scanning is not a permanent evaluation system because the latter generally requires a great deal of time to establish and so cannot meet the time constraints of the formative stages. The permanent evaluation system is the topic of Chapter 8 of the book.

An Example of Comprehensive Scanning. The arthritis self-care project evaluated by Brunk and Goepfinger (1990) employed systematic scanning. The theory-based program plan for the project defined the intervention as a standardized curriculum that taught arthritis self-care behaviors plus problem-solving skills helpful in managing rheumatic disease. The intervention was “packaged” in two distinct modes: home study and small group. Those participants following the home-study mode completed the curriculum in their homes, overseen by trained volunteers called “community coordinators.” Those following the small-group mode met weekly for a total of six sessions, facilitated by trained volunteers called “lay leaders.” Brunk and Goepfinger collected data from several sources, including audio recordings of small-group sessions, weekly informal interviews with caregivers, participants’ records of contact with project staff, completed worksheets, and written communication between project staff and caregivers. The information gained from the multiple methods was wide-ranging and important for detecting problems for program adjustment.

Initially, project designers attempted to identify and recruit community leaders to be the lay leaders. Key informants were asked to nominate leaders, who would then be trained to provide direction to caregivers and conduct the intervention. As it turned out, the evaluators learned, finding enough community leaders to serve nine scattered target areas had been staggeringly difficult. Community leaders named by key informants often were unable to join the project. As a result, volunteers had been recruited and trained to be the lay leaders.

The intervention protocol detailed in the self-care project program plan directed caregivers to present the standardized curriculum and facilitate group discussion. However, the evaluators found that actual service delivery deviated from the plan. They discovered that caregivers had skipped over topics or canceled group discussions. Remedial action taken in the face of these data included emphasizing the standardization of training, censuring caregivers’

behavior when it departed from the prescribed curriculum, and audiotaping class sessions, all in an effort to minimize content variation.

Originally, the self-care project sought to match clients to caregivers in their home communities, maybe even to caregivers known to them. However, with the eventual pool of clients scattered across nine rural counties, that was impractical. In the end, caregivers had been assigned simply to ensure coverage rather than to create the intended pairings of caregivers with community clients.

Brunk and Goeppinger (1990) chose the term *reinvention* to indicate a change in the program plan; the term highlights the positive force such a change constitutes. The researchers, who clearly encountered many changes that had taken place during development of the program, stressed the importance of systematic documentation of any changes in the intervention protocol or other areas. When changes go undocumented, it is later difficult to conduct a high-quality outcome evaluation and interpret its results.

Formative Evaluation Results: Use With Caution

Formative evaluation is useful when it is accepted for what it is. It is a strictly developmentally oriented approach, and its results should be used only for timely fine-tuning purposes. Results of formative evaluation should never be used to describe an implementation's quality, for two reasons. First, formative evaluation aims at providing a "quick fix," mandated by the needs of stakeholders when programmatic problems do surface. A problem identified by an evaluation yesterday may not be a problem today if remedial action has been taken. When conditions are so changeable, it is difficult to make meaningful value judgments. A stable pattern of implementation usually must be in evidence before quality or merit can be judged.

Second, to provide feedback quickly, formative evaluation often must apply research methods elastically, altering certain "prefabricated" methods to suit the circumstances. The small size of survey samples discussed above is one example, and another is found in the interview with key informants. To respect stakeholders' time frame, interviews can include only those key informants who are easily available. The elasticity in application of research methods means that defending the methodology would be difficult if results were used to rate a program's merit. In short, by "stretching" the methodology, formative evaluation invites its designation as a less-than-rigorous approach when measured against traditional research standards.

Still, the value of formative evaluation is not diminished as long as its results are used only for program improvement. The overarching theme of this book is that the different evaluation strategies and approaches are appropriate for different evaluation purposes: Formative evaluation may not be good for merit assessment purposes, but, by the same token, assessment-oriented evaluation may be of little use in program development.

THE PROGRAM REVIEW/DEVELOPMENT MEETING (UNDER THE TROUBLESHOOTING STRATEGY)

In an organizational setting, research is not usually regarded as the sole way to obtain information useful for making decisions about program implementation. Program managers and other stakeholders often rely on organizational meetings to gather information needed to identify problems and propose solutions. Evaluators should be prepared to facilitate such meetings. Adopting the troubleshooting strategy, evaluators called in during the initial implementation stage strive to foster consensus among stakeholders on implementation problems and solutions. One means of deploying this strategy is the *program review/development meeting*. Pressures of time lead to many important decisions about programs being made in ordinary meetings, unaided by the evaluator with his or her empirical field findings. Making decisions in this way is firmly discouraged by the scientific community, with its emphasis on evidence-based choices. Whatever its weaknesses, however, for the foreseeable future, the meeting convened to discuss issues and make decisions will continue as a modus operandi within most organizations. Interest in program development is growing among organizations, though, so the time may be ripe for evaluators to examine the meeting-based approach to decision making. Perhaps, if pursued with caution and a recognition of the limitations involved, this approach could become another option for evaluation.

The purpose of the program review/development meeting is to have program supervisors and implementers (or their representatives) gather, in the presence of evaluators, to talk over challenges facing their program. This approach requires the evaluator to serve as discussion facilitator *and* consultant, systematically reviewing with the stakeholders the major difficulties with the program's implementation and proposing problem-solving strategies. Discretion should certainly be exercised in the decision to use the program review/development meeting approach, which largely generates information appropriate for internal audiences only. It does not have the capacity to meet

accountability requirements. Stakeholders, especially funding agencies, are strongly urged not to use the program review/development meeting as a substitute for other necessary kinds of evaluations. Evaluators must communicate the limitations of this approach to stakeholders. A discussion of guidelines for using the program review/development meeting approach follows.

Program Review/Development Meeting Principles and Procedures

It is important to distinguish a program development meeting from the regular staff meetings convened by an organization. Regular staff meetings are usually conducted and controlled by a supervisor, and it is here that the limitations of such meetings start. The supervisor, having authority over the implementers, is perhaps not the first person to whom they would voice their observations about problems. Doing so may seem too much like an acknowledgment of incompetence. Furthermore, supervisors may lack expertise in steering the discussion to systematically explore implementers' views. The risk of incomplete discussions—because of the supervisor's presence and because of the supervisor's possible deficiencies—threatens the meeting group's identification of problems and remedial actions. The program review/development meeting tries to overcome the limitations of regular staff meetings by including evaluators from outside the organization as facilitators and consultants. In a program review/development meeting, supervisors become equal partners with staff members, and the evaluator steers the discussion. Evaluators have much to contribute to such meetings.

In the role of facilitator, an evaluator helps create an open, safe atmosphere in which participants freely express their opinions and recount their experiences with implementing the program. For this to happen, the evaluator should be external (independent from the organization and having no stake in the program) rather than internal (regularly employed by the organization). In the role of consultant, an evaluator uses knowledge of evaluation (such as the framework of program theory) to systematically set an agenda, provide background information, and present all important issues before the meeting. The evaluator can also provide, as needed, a consultant's input concerning options to resolve problems.

Before the program review/development meeting, the evaluator should transmit to each supervisor and staff member a draft agenda intended for review and comment. Ahead of the meeting date, the evaluator should also

secure from the program director and other supervisors a commitment to ensure a safe environment for discussion. The program review/development meeting should start with an announcement of the meeting's purpose and the setting of ground rules for discussion. General ground rules should be that individual opinions are honored and information from the meeting is destined for use in improving programs rather than punishing people. Again, supervisors are to be regarded as equal partners in this setting. The conceptual framework of program theory, especially the action model, is available as a guide to evaluators as they facilitate systematic discussions of the major areas of implementation and the problems therein. As remedial strategies begin to be developed in the meeting, the evaluator offers professional opinions for consideration.

Facilitating a program review/development meeting takes excellent communication and facilitation skills. Not every evaluator is constitutionally suited to the task, but disinclination or lack of skill here can be overcome by teaming the evaluator with a professional facilitator. The two work closely together to prepare the agenda and materials and serve as consultants at the meeting. The need for external evaluators, pointed out above, is strongest when stakeholders have highly divergent interests and backgrounds. In the midst of competing interests, the external evaluator tends to strike staff members as neutral and credible. (Large-scale programs, especially, benefit from using external evaluators.) Under the following very specific conditions, it may be possible to have internal evaluators facilitate a development meeting:

- Good working relationships exist among supervisors and staff.
- The internal evaluators are very knowledgeable about program evaluation.
- The internal evaluators have good facilitation skills.

Internal evaluators certainly have one advantage over external evaluators: Their services are low cost or even no cost when supervisors agree to that arrangement.

Program Review/Development Meeting Advantages and Disadvantages

The program review/development meeting has several advantages when it comes to providing information to enhance programs in the initial implementation stage.

- Implementers are given a sense of ownership and may enthusiastically buy in to the problem identification and solution process. The meeting is an opportunity to express views and concerns about implementation and needed action. Implementers recognize that they are a real force in development, which may mean they are likelier to support proposed remedial action and other changes.
- Costs remain low. Obtaining evaluators to facilitate meetings requires some money, but much less than most other evaluation approaches requiring data collection in the field.
- Program review/development meetings produce feedback that can be turned around quickly. To write a summary report of what was learned in a development meeting takes just a few days to a few weeks, depending on what the evaluator and organization have arranged.

However, this approach also boasts a handful of significant *disadvantages*:

- Input at a development meeting can be quite impressionistic. Discussions consist largely of implementers' impressions; the accuracy or validity of impressions is not checked or verified. For example, if an implementer says that some clients are reportedly swapping program-provided food coupons for street drugs, discussion of the issue can be intense, as you might imagine, but nonetheless present no factual information about the alleged problem. Additionally, implementation problems are usually multifaceted, whereas each individual implementer views problems from a single, personal perspective. It is not uncommon for implementers' input about problems to be partial or fragmented. Action steps coming out of the meeting then may be inappropriate or ineffective.
- The development meeting approach emphasizes a consensus-building process, not necessarily accuracy. To encourage participation and satisfy the meeting members, the facilitator seeks parity, representativeness, and inclusiveness. Although this certainly fosters consensus, it has little to do with accurately identifying problems and their optimal solutions.
- Vocal or articulate implementers may dominate a program review/development meeting. As in most meetings, some participants in program review/development meetings will be more stirred up, more articulate, or just more comfortable speaking in public than other participants. Facilitators work to encourage those on the low end of the spectrum, but, despite their efforts, some meetings are dominated by the outspoken and the well-spoken (squeaky wheels do get the grease). Even worse, at times, the more vocal and articulate individuals bring personal agendas to a meeting and manipulate the proceedings to serve their own ends.

Example of a Program Review/Development Meeting

Gowdy and Freeman (1993) convened a development meeting as an analysis and development tool for a program helping low-income women become and remain employed to achieve economic self-sufficiency. Services provided by this program included GED instruction, a job readiness course, employment development and placement services, and support services (transportation and child care assistance, a clothing bank, counseling, referral, and follow-up). Most clients were single mothers in their late 20s or early 30s. Gowdy and Freeman used a conceptual framework they called the *program model* (which is similar to the program theory discussed in this book) to facilitate the program development meeting. Internal evaluators were chosen because the conditions cited above for their use had been met. The evaluators began by reviewing program documents, with the program model as guide. A memorandum and copies of the program model were sent to participants in advance of the meeting, which was held at the agency and kept deliberately informal. A 3-hour meeting was planned.

Evaluators' tasks during the meeting included "translation" of the program model's meaning, purpose, and operation into the staff's operating language. Evaluators were responsible for keeping the meeting on track, as well, and creating a safe, open discussion in which all could participate. They met with success in terms of valuing each individual's opinion and avoiding domination of the meeting by any one person. As Gowdy and Freeman (1993) put it, "The receptionist's experiences of the program were listened to and considered as much as those of the director and direct service staff" (p. 69). Important findings came from this meeting pertaining to six aspects of the program.

Implementing Organization and Implementers. Meeting participants indicated that the program staff were predominantly African-American and exclusively female, but this was consistent with the clients served. Nevertheless, hiring a Hispanic woman was suggested to increase diversity. Discussion established that staff relationships were good, but there was a need to improve attendance at and communication during weekly staff meetings. Discussion also established that meeting participants believed the training of program staff was inadequate; they felt a strong need for more orientation activities for new staff members, monthly in-service training, funding to attend conferences and seminars, and tuition reimbursement. Furthermore, it was indicated that the program's resources were not sufficient to support beneficial home and workplace visitation.

Target Population. Opinions expressed at the meeting consistently spoke to the program's success in targeting and providing services to vulnerable, low-income

African-American women. The group believed the services should and could be expanded to benefit other vulnerable minority women, such as Hispanic and Asian women. As for potential barriers to participation, the group acknowledged that the agency was centrally located and accessible by public and private transportation. From their perspective, enrollment of clients had been no problem, thanks to active outreach and effective recruitment strategies, a flexible intake schedule, and transportation and child care assistance.

Service Delivery. In general, service delivery was adequate, according to meeting participants. If there was a weak area, they felt, it was that little systematic attention was paid to clients during the postgraduation period. Despite the establishment of a goal to support women for up to 1 year while they entered the job market, no clear guidance facilitated ongoing intervention once women completed classes and sought employment. Clients experienced frustration when a job did not materialize immediately, participants reported. Clients also experienced other strong emotions during this period: anxiety, excitement about learning, fear of success, and gratitude for help received. Meeting participants indicated their belief that a good deal more attention needed to be paid to clients' emotions as part of the change process.

Ecological Context. According to those present at the development meeting, the agency had a positive reputation as an advocate for minority women.

Goals and Outcomes. Participants said that goals for helping the low-income women reach economic self-sufficiency were explicit, clear, and focused on desired outcomes. They did suggest, though, that process goals, such as acquiring job-related skills, might enhance the program.

Determinants. The participants indicated that precisely how the services of the program altered the relationship of women and poverty had never been clearly understood. They identified empowerment as the determinant for achieving the goals.

Gowdy and Freeman (1993) reported that this development meeting led to significant modifications of the program. As an example, changes were made to deal with inadequacies of postgraduation follow-up. A peer support network was designed for the graduates that, in addition to providing mutual support, fostered role-modeling and skill development. Additional staff were brought on board, strengthening the program's case management capacity. Staff started to touch base and maintain regular contact with clients via mailed surveys and

personal outreach. Finally, a committee was named to review and assess progress in light of the changes.

This example demonstrates that information from the program review/development meeting is useful to internal audiences *if the meeting has been approached with preparation and managed expectations*. The information, however, cannot be used as evidence of a program's quality or effectiveness for an external audience because the approach cannot ensure the information's credibility.

BILATERAL EMPOWERMENT EVALUATION (UNDER THE DEVELOPMENT PARTNERSHIP STRATEGY)

Evaluation activities are ever more diverse, and program stakeholders have started inviting program evaluators to collaborate as their partners in the development (design and implementation) of programs. When such a development partnership is formed, evaluators become members of the program development team and part of the decision-making process. The typology refers to this as the *development partnership strategy*. Evaluators using this strategy will negotiate points with key stakeholders, but, more importantly, they will be empowered to put evaluation results to work to devise or strengthen programs. The latter is not by any means an easy job, as it requires sustained effort from evaluators, but the development partnership strategy is one available tool.

An evaluation approach associated with the development partnership strategy is the *bilateral empowerment evaluation*. Bilateral empowerment results when evaluators welcome stakeholders to join the evaluation process and stakeholders welcome evaluators to join the program development process. Bilateral empowerment evaluation has gained strong momentum in the evaluation of community coalition programs. Whether such an organization is termed an "alliance," "consortium," "partnership," or "network," its general aim is to empower a community by building its capacity to solve community problems. Funds for community coalition programs frequently lack well-specified requirements, allowing the community freedom to develop strategies and take action as it sees fit. Furthermore, the goals of programs like these tend to be broad and ambitious; they might call for reducing drug abuse or ending discrimination, for instance, throughout an entire community. Evaluating community coalition programs is necessarily a complex and dynamic process because the work proceeds on two fronts—the evaluation process and the program development process.

Evaluation Process

The evaluative procedures of the bilateral empowerment evaluation encourage sharing. The evaluation process is a mutual learning experience for stakeholders and evaluators, and each has empowered the other to improve the program. Evaluators also are obligated to give technical assistance to stakeholders and to build their capacity to weigh important issues wisely. In the bilateral empowerment relationship, joint decision making is the norm, extending to tasks such as conceptualizing community development and devising instruments to measure it or gathering pertinent data.

Example of Bilateral Empowerment Evaluation

A good illustration of bilateral empowerment evaluation is found in the evaluation of a community prevention alliance conducted by Goodman et al. (1996). The alliance proposed to work for change in a community's norms regarding the use of tobacco products, alcohol, and street drugs. It was hoped that the organization's efforts would (a) alleviate problems in the community's workplaces concerning use of these items, (b) reduce addiction and violence among 12- to 17-year-olds in the community, and (c) limit the annual incidence of HIV/AIDS and other sexually transmitted diseases in the community. The stakeholders had worked closely with evaluators to plan, develop, and implement the intervention program. Goodman and colleagues broke down the process into three "phases" of work, as follows.

Phase I: Program Formation. The very first need was to create the evaluation coalition by hiring staff and recruiting members from across the community. Then the coalition was involved in carrying out a needs assessment, data from which would inform the determination of its intervention strategies. In this phase, evaluators serving in the coalition assisted stakeholders with completing a logic model of the intervention and developing indicators to measure the progress of program development. The evaluators attended coalition meetings to debrief the membership and discuss the latest evaluation results. They continually monitored the work on the program plan and encouraged stakeholders to fine-tune both systematically and as needed to ensure that the implementation would ultimately remain consistent with the plan. For instance, they monitored the effectiveness of coalition meetings because the organization of the new coalition depended on that effectiveness. When attendance and participation at meetings were both found to be lacking, evaluators brought the

relevant data to the coalition members for discussion, leading to a strategy to improve rates of attendance and participation.

Phase II: Plan Implementation. Eventually, the time came to implement the strategies approved by the coalition: awareness campaigns, service programs, and policy initiatives. Evaluators began to monitor the level of effort put forth by the coalition in pursuing each of these. They found evidence of a strong emphasis on awareness campaigns, one that detracted from the implementation of other activities. In addition, evaluators felt that awareness campaigns were not likely to create lasting change. Evaluators turned to the evaluative data to interest coalition members in refining the intervention to better stimulate the community to change. The evaluators also worked to build the coalition's capacity to implement the program plan effectively and to increase its ability to use evaluation results well.

Phase III: Impact. The final phase featured efforts to institutionalize coalition strategies, determine what the community-wide results of the intervention were, and preserve the coalition following termination of the grant. A successfully implemented coalition could, it was thought, boost community awareness of, concern for, and action on substance abuse, violence, HIV/AIDS and other sexually transmitted diseases (STDs), and teen pregnancy. Goodman et al. (1996) suggested using such research methods as a survey of key community leaders, a broader community survey, and trend analysis for use in the assessment; however, performing the impact assessment was beyond the scope of their study.

The Evaluators' Role

The essence of the evaluator's role in program planning and development under the bilateral empowerment evaluation approach is expressed well in Goodman et al. (1996):

In general, we stress to Alliance members that the best ways in which we [evaluators] can be helpful are by being dedicated to an ongoing relationship with the coalition and by providing honest feedback that is based on data, open sharing of information, problem solving, negotiation, good will, and support of the coalition's effort. While the coalitions that we evaluate do not always agree with our approaches, conclusions, or recommendations, they view us as valued members who provide important feedback.

Without earning the trust of our community coalition through open communication, negotiation, and compromise, we do not believe that our assessment approach is feasible. (p. 58)

Pros and Cons of Bilateral Empowerment Evaluation

In a fluid, complex program such as a community coalition or consortium, evaluation information is usually recognized as crucial for directing the program's development. Stakeholders with this mind-set tend to invite evaluators to partner with them in the planning and program development process. This participatory quality is seen again in bilateral empowerment evaluation, in which evaluators move beyond formative evaluation to serve as planning and development team members. They work closely with stakeholders to select an intervention, structure the program, and solve implementation problems.

Bilateral empowerment evaluation, like any other approach or strategy, has its pros and cons. Its chief advantage is that it maximizes the impact of evaluation data on the decision-making process. Bilateral empowerment allows evaluators to apply the development-oriented process evaluation discussed in this chapter. Later on, however, these evaluators are likely to need to bow out of assessment-oriented evaluation concerning the merit of the program they helped launch. Because of the risk of conflict of interest, the credibility of an outcome evaluation carried out by evaluators using the bilateral empowerment strategy would be compromised.

QUESTIONS FOR REFLECTION

1. If you are a program director, what are your main concerns when your program is in the initial implementation stage? Give examples.
2. Is it appropriate to collect information regarding implementation problems and management during the program review/development meeting? Explain.
3. Troubleshooting is the process by which major problems with the program are discussed. Why is it important to include troubleshooting in the initial implementation stage?
4. Discuss the steps of formative evaluation application. Explain why each step is significant.
5. Compare and contrast formative evaluation and formative research as discussed in this and previous chapters. Give examples.

6. Give real-life examples of each of the four types of formative evaluation.
7. In your opinion, which formative evaluation example provides the evaluator with the most pertinent information? Is it important to utilize all six steps?
8. Why is it necessary to utilize formative evaluation during the “fine-tuning” stages of the program?
9. Discuss what could potentially happen if an evaluation was not conducted within the stakeholder’s time frame.
10. Discuss the principles and procedures of a program review/development meeting.
11. Discuss the advantages and disadvantages of program review/development meetings.
12. Explain how the disadvantages of program review/development meeting can be mitigated (e.g., How would you ensure every participant is heard if you have a vocally dominant participant?).
13. What role does an evaluator play in a program review/development meeting?
14. Discuss the strengths and weaknesses of bilateral empowerment evaluations.

CHAPTER 7



Assessing Implementation in the Mature Implementation Stage

Programs in the mature implementation stage are those in which the procedures and rules of implementation have become routine. This chapter will discuss the use of *process evaluation to assess how a program is being implemented*.

As discussed in Chapter 1 and the comprehensive evaluation typology in Chapter 2, there are three kinds of process evaluation: constructive, conclusive, and hybrid. At the mature stage, stakeholders could ask evaluators to conduct an evaluation to identify implementation problems in a timely manner, as discussed in the previous chapter. This kind of evaluation is called a constructive process evaluation. More often, however, stakeholders ask evaluators to assess how well the program was implemented; this evaluation is called a conclusive process evaluation. Alternatively, stakeholders may ask evaluators to conduct a hybrid process evaluation that serves both conclusive and constructive purposes. These three types of process evaluation that are conducted at a program's mature stage are discussed as follows:

Constructive Process Evaluation. Even a mature program may be subject to problems of a programmatic nature. When this happens, stakeholders may want evaluators to assess the problem quickly and provide information to help them develop a remedial strategy. This kind of evaluation is called a constructive process evaluation. Thus, those development strategies and approaches previously introduced that stress timeliness (see Chapters 4–6) can also be used in evaluation at the mature implementation stage. These strategies can be

just the tools needed to identify stakeholders' development needs and pinpoint the corresponding evaluation techniques.

Conclusive Process Evaluation. At the mature implementation stage, concern for accountability begins to grow among external stakeholders, such as funding agencies and decision makers. Internal stakeholders, too, become inquisitive about accountability. A conclusive process evaluation assesses how well the program has been implemented.

Hybrid Process Evaluation. Stakeholders, especially program directors and staff, are particularly interested in information on how to better serve clients. They hope process evaluation can provide information on not only the quality of implementation but also exact areas that need improvement and how to improve. An evaluation that serves both conclusive and constructive purposes is called a hybrid process evaluation.

This chapter starts with a brief review of constructive process evaluation before moving to a discussion of conclusive process evaluation and hybrid process evaluation.

CONSTRUCTIVE PROCESS EVALUATION AND ITS APPLICATION

Constructive process evaluation can be very productive in the mature implementation stage when a program plan (program scope and/or action plan) needs to be revisited for clarification or modification or when implementation problems call for some troubleshooting.

Modifying or Clarifying a Program Scope and Action Plan

At times in a long-running program, stakeholders such as the program director, implementers, and decision makers can come to feel that they have lost sight of the goals being pursued or the direction in which the program is moving. They want to revisit their program scope and action plan to clarify or even reshape the program. Development-oriented strategies are useful here, and an especially popular one is *development facilitation*. Many consensus-building tools are available within program evaluation that can facilitate the work of stakeholders needing to redefine their goals or reprioritize a set of existing goals. When prioritizing goals is the main interest, traditional techniques—for example, the

Delphi method, nominal group technique, and multi-attribute utility method—can serve the purpose (see discussions in Chen, 1990). However, when evaluators and stakeholders are pursuing issues that affect relationships among program objectives and goals, they will find more helpful the *conceptualization facilitation* and *concept mapping* approaches (see the typology, Table 2.1). Both of these techniques are related to program theory, but each offers a unique advantage. Conceptualization facilitation is appropriate when relationships among components of the rationale and plan need to be identified or made clearer (as elaborated in detail in Chapters 4 and 5); concept mapping, a quantitative tool, is especially helpful when a program has numerous goals and activities.

Conceptualization Facilitation Approach

Facilitating stakeholders' efforts to elucidate their program rationale and program plan requires building consensus, so the working group format of the conceptualization facilitation approach is preferred to intensive interviewing. The conceptualization facilitation approach works systematically, so it can be employed to develop rationales and plans as well as improve them. The evaluator engaged to help with either task will find in Chapters 4 and 5 much guidance for formulating evaluation designs and procedures of a conceptualization facilitation type.

Concept Mapping Approach

Concept mapping is a quantitative tool with which stakeholders can clarify program objectives and identify relationships among them (Trochim & Cook, 1992). The approach is a structured process for soliciting the opinions of representatives of various stakeholder groups. Ideas for the program contributed by these representatives are organized in a graphic representation called a *concept map*. The concept map, then, illustrates the stakeholders' theory of the program's basic components and their interrelationships. Procedures for applying concept mapping are illustrated in the following example.

An Example of Concept Mapping. A YMCA drop-in center studied by Mercier, Piat, Peladeau, and Dagenais (2000) stated its purpose as offering 10- to 17-year-olds an informal setting for unstructured and structured activities after school hours and on Saturdays. Structured activities included sports and recreational programs, educational and sensitization programs, and informal counseling and referral services. This YMCA center saw itself as an alternative to unsupervised settings and a deterrent to undesirable behavior such as substance use, intergenerational and ethnic conflict, sexual activity and resulting pregnancy, and dropping

out of school. The board members and staff of the drop-in center had expressed a need to better understand the preventive function of the center to make their strategic planning more effective.

The center's director and three staff members participated actively in devising a concept map. At their initial meeting with evaluators, they were asked the general question "In what way can/does the youth center contribute to prevention?" They brainstormed answers in the form of short statements, and a total of 98 statements were generated. In a subsequent meeting, each participant was asked to rank the importance of each statement relative to the others and to sort the statements into thematic groups. This yielded, for the concept map, the basic groupings or components of the program. At a follow-up meeting, participants were asked to identify causal relationships among these components. Data analysis was then conducted, and staff identified three desired outcomes for the program: (a) to offer youth an alternative to the street, school, or family; (b) to promote personal and social development; and (c) to sustain leadership development. Staff members believed that these outcomes would be most effectively pursued through such means as the flexibility of the center program and activities, special events, the freedom to experiment while supported by supervision, recognition of achievement leading to a sense of self-worth, and so on.

Troubleshooting Implementation Problems

Stakeholders may not always be satisfied with an implementation process, even when their program has reached the mature implementation stage. They may invite evaluators to look into perceived problem areas and generate relevant information helpful to them as they consider their next move. In such cases, troubleshooting strategies and approaches (discussed thoroughly in Chapter 6) should meet stakeholders' needs.

A study of a health education program for pregnant women in Chile (Foster, 1973) provides an interesting example of constructive process evaluation at the mature implementation stage. Public health centers there arranged for women who had just been informed that they were pregnant to attend prenatal classes. The program was modeled on an American program quite popular and successful in the United States. In Chile, however, its success had been limited, chiefly because expectant mothers failed to attend classes. Research showed a problem with an aspect of the program's service delivery mode—its setting. The targeted women objected to being taught in classrooms like children. Prompted by this finding, public health decision makers restructured the classes as "clubs," which typically met in the women's various homes. In Chile, club membership connotes

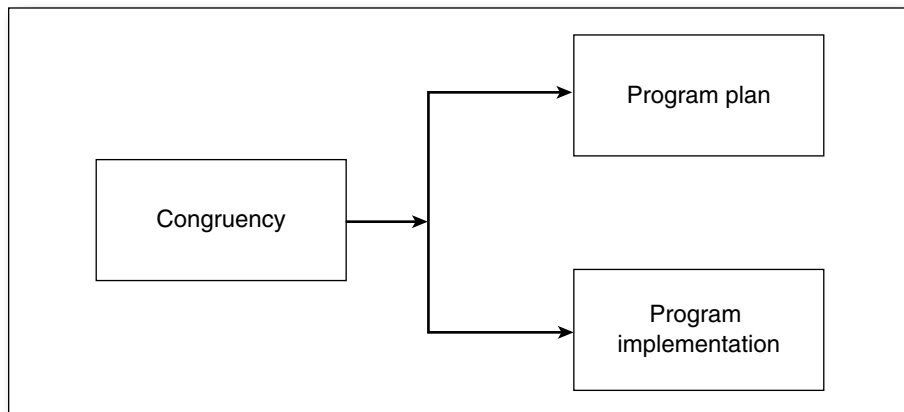
middle- to upper-class status, so the women then enjoyed participating. The program contributed refreshments for each meeting, and the intervention became a social affair where conversation was mixed with exchanges about prenatal care. Changing the service delivery mode led to a quite successful program.

CONCLUSIVE PROCESS EVALUATION AND ITS APPLICATIONS

Constructive process evaluation is conducted primarily for internal use to address problems in a program immediately (see Chapter 5). The results are not based on rigorous methodology and do not provide comprehensive information for determining the merits of program implementation. When external stakeholders (e.g., funding agencies, decision makers) and/or internal stakeholders want to know how well a program is being implemented, a conclusive process evaluation is called for. The conceptual framework of a conclusive process evaluation is illustrated in Figure 7.1, which indicates that conclusive process evaluation assesses the actual implementation process vis-à-vis its original program plan.

There is a tendency to think of accountability as meeting (or falling short of) program goals. In actuality, program outcomes comprise just one area of accountability. Another area of accountability, which is of considerable importance, is program implementation. Funding agencies, decision makers, and other external stakeholders are very interested in process questions such as, *Whom is the program, as implemented, serving?* and, *What is the quality of the services being provided?* These and similar questions can be answered through

Figure 7.1 Conceptual Framework of Conclusive Process Evaluation



process evaluation. The results of process evaluation will also, of course, be of interest to internal stakeholders (e.g., program directors, implementers), who have much invested in the quality of implementation.

Furthermore, conclusive outcome evaluation is also useful for the interpretation of program outcomes. An *outcome evaluation*, which will be discussed later in this book, may tell whether or not an intervention has indeed effectuated an outcome. However, evaluation findings concerning the relationship of intervention to outcome per se are difficult to interpret or use in the absence of information about the implementation. For example, what is signified when outcome evaluation shows no relationship between intervention and outcome? Is it that the intervention is inappropriate? Or were there serious flaws in the implementation process? The data from conclusive process evaluation generate the contextual information necessary for useful interpretation of outcome evaluation results. For example, if a program produces none of its expected outcomes, conclusive process evaluation will be able to provide information about the source of the failure: Is it due to erroneous conceptualization of the program, or is the cause some deficiency in the implementation?

How to Design a Conclusive Process Evaluation That Fits Stakeholders' Needs

As indicated in Figure 7.1, a common tactic of a conclusive process evaluation is to conceive of the quality of a program's components as a congruency between stakeholders' intentions for the program (their program plan) and the facts of the program's implementation. Congruency between program plan and program implementation is widely understood to signify a high-quality implementation.

In designing a conclusive process evaluation, evaluators always need, first and foremost, to understand clearly stakeholders' needs as well as the evaluation strategy approaches most appropriate for meeting those needs. These two tasks are discussed in the following sections.

Clarifying Program Intentions

The evaluator should have a clear idea of the program's intentions before choosing a conclusive process evaluation strategy. When the intentions are plainly set forth in a well-developed program plan, this evaluation step is straightforward. Clarification of intent is accomplished mainly by discussing the program plan with stakeholders, ensuring that it is up-to-date and reflects their views. However, if there is no sound, formal program plan in place, the evaluator

needs to compile and consult with whatever documents do exist and meet at greater length with key stakeholders. Performance assessment emphasizes target groups and intervention components. Usually, clarification of the proposed intervention's intent and the target groups to be reached can be achieved by holding one or two meetings with key stakeholders. If the stakeholders are interested in an enlightenment assessment, however, a few meetings are likely to be necessary to identify the major components of the program. Chapters 5 and 6 present principles and procedures for creating a program plan, and these make a very useful guide to identifying these major components. (Note that involving the evaluator as a facilitator during the development of the program plan reduces the amount of time necessary for clarifying intentions.)

Occasionally during work to clarify stakeholders' intentions for their program, the evaluator runs into factions at odds over certain program components. Imagine, for example, an HIV counseling and testing intervention in which program managers and officers at the federal level intend to restrict services to high-risk people but implementers at the state level intend to offer some response to anybody who walks into their clinics. Ideally, consensus about a program's intentions precedes evaluation, but if only a few points are disputed, the evaluation can proceed. In fact, the evaluator can use the evaluation as an opportunity to collect empirical data—findings from the field—that may allow the opposing stakeholders to resolve their differences. In this example, both federal HIV program officials and state HIV implementers would, perhaps, appreciate knowing that data show overly restrictive clinics discourage both high- and low-risk people from visiting.

Selecting Evaluation Approaches

Conclusive process evaluation attempts to provide credible information about the quality of implementation. Naturally, such an exacting review requires time—a good deal more than constructive process evaluation requires. Because emphasis is *not* on the speedy return of feedback, the methodology of conclusive process evaluation has to be rigorous, and the evidence has to be convincing. Conclusive process evaluation is conducted through fidelity assessment. The next section will discuss various fidelity approaches from which evaluators and stakeholders can select.

APPROACHES OF CONCLUSIVE PROCESS EVALUATION

The *fidelity evaluation approach* is the major approach associated with conclusive process evaluation in the mature implementation stage. Fidelity evaluation

gauges the degree of congruency between intervention and target groups as planned and intervention and target groups as implemented. Many fidelity evaluations (as shown below) pursue one issue: to assess whatever element of the action model conceptual framework is of special interest to stakeholders. Any of the four popular types of fidelity evaluation about to be introduced here can serve that need, depending on which element of the model will be scrutinized. The four types are intervention fidelity evaluation, referral fidelity evaluation, service delivery fidelity evaluation, and target population fidelity evaluation. Chen (1990) covered several other types of fidelity evaluation associated with the remaining program components, including contextual support and organizational collaboration, and provided relevant evaluation strategies and examples.

Intervention Fidelity Evaluation

Evaluators have used the terms *intervention fidelity* and *treatment integrity* to refer to what in this book is called *intervention fidelity evaluation*. The point of intervention fidelity evaluation is to see whether an intervention implemented in the field is turning out as patterned by the original program plan. Generally speaking, implementations that reflect most clearly the intent expressed in the program plan are those of higher quality and thus those most likely to work.

To conduct intervention fidelity evaluation, stakeholders and evaluators must first identify the crucial elements and strengths of the intervention as it was *intended* to function. An intervention often consists of a number of elements. An antibullying intervention in a school system, for example, might include the following elements enumerated in the associated program plan: adopt school policies on bullying; hold training and discussion sessions for students, staff, and parents; obtain additional play equipment; and utilize adult supervision. Evaluators need to work with stakeholders to identify the *crucial* elements of the intervention so that they can be included in the assessment.

An intervention's fidelity to a program plan can be assessed using one or more of the following measures: coverage, strength, and intensity. Measuring *coverage* means asking whether the real-life implementation covers all crucial activities of the intervention as prescribed during planning. Measuring *strength* means determining whether the implementation includes as much of the intervention, per session, as planners intended. That is, is the prescribed "dose" of the intervention being administered at one time? Measuring *intensity* means counting the number of sessions, or times the intervention was carried out, to see whether the prescribed number or dose was provided. For instance, training sessions within the school's antibullying program could be assessed in terms of

whether they covered adequately the prescribed antibullying topics, the number of minutes they lasted, or the total number of sessions offered to students.

Example of an Intervention Fidelity Evaluation. An intervention fidelity evaluation was conducted on a school-based nutrition education program in Georgia (Davis et al., 2000). The program was established to help fourth- and fifth-grade students consume more fruit, fruit juice, and vegetables (nick-named “FJV”) each day. The intervention in this program had three major elements: its curriculum, family activities, and point-of-purchase activities. The curriculum was delivered in 12 sessions, each lasting 40 to 50 minutes. Teachers presented the curriculum, which comprised various tasks classed in categories such as affect (increasing students’ enjoyment of eating FJV) and asking skills (enhancing students’ ability to ask for FJV at home). Researchers observed classrooms to gather data about the fidelity of the curriculum’s implementation. Parents and families of the students were also drawn into the program via homework assignments and videotapes about FJV. To gather data about the fidelity of implementation of FJV family activities, evaluators used telephone interviews with parents to ascertain whether and how many homework assignments and videotapes had come home and to ask whether the family had joined in homework activities, viewed videotapes, and/or set FJV goals. During these telephone interviews, parents were also asked about participation in point-of-purchase activities held in the evenings at local grocery stores. The additional questions, plus evaluators’ observation of the in-store sessions, generated the fidelity data concerning this third element of the implementation.

Data analysis indicated that, overall, elements of the program had not been implemented as designed. Teachers had failed to deliver the entire curriculum, selectively underimplementing messages that were crucial to promoting behavioral change. Few families had attended evening point-of-purchase sessions. At-home family participation was modest, and it declined substantially between the fourth grade and the fifth. More than one third of the parents of fifth graders said they did not participate in any of the child’s FJV homework activities; about the same number said they had not received a videotape.

Referral Fidelity Evaluation

The fidelity evaluation approach is also used to assess the adequacy of the *referral process* of a program. This kind of evaluation is particularly beneficial to programs serving populations whose behavior is high risk, such as drug

abusers. Clients enrolling in an intervention for high-risk behavior usually have multiple problems. For example, a person admitted to a mental health treatment program may be dealing with alcoholism, homelessness, hunger, and more. For this person, the success of the mental health treatment depends not just on the intervention but also on the alleviation of other problems. Thus, a referral network of other programs that serve the individual's additional needs is vital to the success of the mental health intervention. The mental health program could easily fail unless it recognizes related barriers to the client's well-being and knows who can help lower these barriers. The chief question asked by referral fidelity evaluation is "Is there a referral process in this program, and does it function as intended?"

Example of a Referral Fidelity Evaluation. An example of evaluating referral services for program clients is found in the work of Marx, Hirozawa, Chu, Bolan, and Katz (1999), who wrote about clients needing referrals from an HIV-testing/counseling program to prevention services. The program, known as Counseling, Testing, Referral, and Partner Notification (CTRPN), is the largest standardized HIV prevention effort in California, but it makes only brief contact with clients, generally encompassing only two sessions. To serve clients effectively, CTRPN needed to provide further attention to high-risk individuals by referring them for additional prevention services. Its guidelines clearly indicated that one of CTRPN's major objectives was to provide referrals to HIV-positive and high-risk HIV-negative persons for necessary medical, preventive, and psychosocial services.

Data for the study of CTRPN's referral fidelity included information from the San Francisco Department of Public Health and the city's STD clinic. The research showed that the overall referral rate of referral by CTRPN was low: A referral had been received by 19.1% of the health department sample and by 10.6% of the clinic sample. The authors concluded that opportunities were being missed to link high-risk clients who had been tested or counseled to additional HIV prevention services outside CTRPN. They urged that the referral component of the HIV counseling/testing program be improved.

Service Delivery Fidelity Evaluation

Service delivery fidelity evaluation ascertains congruency among the setting, mode, and procedure of service delivery as planned and as actually manifested. Service delivery is one part of an intervention that inarguably should be assessed. Service delivery fidelity is especially crucial for intensive programs like those

often found in the fields of mental health, alcoholism, and substance abuse. Intensive programs often use sequential steps or sessions to complete service delivery; serving clients properly involves steps such as intake, screening, risk assessment, case management, client readiness, treatment sessions, and so on. The success of a later step depends on how appropriately the earlier step or steps were implemented, as seen in this hypothetical example of an alcoholism treatment center. When admitted to a treatment program, the alcohol-addicted individual must be immediately assessed as to whether detoxification will be needed before treatment begins. If detoxification is ruled out, the next step is for staff to determine the treatment that is most appropriate. Appropriate screening and diagnosis early on help to ensure that clients receive an effective treatment in a timely manner. Appropriate procedures for screening and diagnosing newly admitted clients are usually well documented in a program's operating manuals and other guidelines. Service delivery fidelity evaluation can be used to explore whether an intervention is being implemented in keeping with all prescribed procedures and rules.

Example of Service Delivery Fidelity Evaluation. Further illustrating service delivery fidelity evaluation is a Taiwanese project that sought to bring middle-school dropouts back to the classroom. One intervention used by the project was teacher counseling of dropouts and their parents together, encouraging the children's return to school. The counseling was delivered via regularly scheduled home visits. In 1993, the author participated in a process evaluation of the project, through which the stakeholders and the author eventually learned that teachers were making very few of the prescribed home visits. The collected data showed that meeting with families was a much harder task than program designers had imagined. Not only did busy working parents have little time for meeting teachers during the day, but teachers (most of whom were female) did not feel safe visiting distant—and often crime-ridden—neighborhoods at night. Beyond such logistical stumbling blocks, many parents also felt that they lacked much control over their children's behavior; they felt they could not, for instance, guarantee their children's presence during home visits. The completed service delivery fidelity evaluation indicated that the home visits were failing as a service delivery mode for this project.

Target Population Fidelity Evaluation

Target population fidelity evaluation looks at programs' contact with their target populations. Programs must reach sufficient numbers of clients from the

specified target population in order to be effective. It might seem safe to assume that any implementing organization would maintain records of this important information, but many, unfortunately, lack good systems to manage client data. This is especially true of programs working with street outreach, capacity building, and similar techniques. Street outreach service programs almost always have information about how many hours outreach workers spend on the streets in a day, and where. However, they seldom gather information on how many contacts were made or who enrolled in the program as a client. Similarly, a program providing capacity-building services to community organizations may maintain information on how many calls are fielded per day yet have no data on which organizations called, the nature of the calls, or the services ultimately provided. An evaluator who has the opportunity to become involved in early stages of program development may be able to facilitate the establishment of a data collection system for recording client information. If the evaluator was not involved in this way, however, and if implementation began long ago without a data collection system, other means for estimating how many target population members are served are available. In the case of the street outreach program, for instance, a survey of a representative sample of residents in the targeted areas could be used to estimate how many people are being contacted by the outreach workers. In the capacity-building case, evaluators could contact those community organizations that have telephoned the program over a specific period and talk with them to reconstruct the purposes of the calls and the services provided. To conduct a target population fidelity evaluation, evaluators need to ask three main questions:

1. *How many clients were served by the program during a specific period?*

Counting the clients served may or may not be a straightforward prospect. The evaluator must remain aware of the distinction between clients recruited and clients served, although this distinction is not much of an issue for programs delivering their services without delay. Through street outreach, services can be provided on the spot and clients counted straightforwardly. The same is true of a hotline program: Count the number of people who make calls that get answered. In such cases, the intervention follows on the heels of recruitment and is soon completed. With intensive programs, however, as discussed above, weeks or even months may be needed to complete services, and clients must be tracked over that period in order to be counted accurately.

2. *How many of the clients served come from the target population?* It is not necessarily true that all those served come from the intended target population. Implementers tend to want to serve clients who are easily accessible. As

an example, consider a program intended to entice school dropouts back to the campus. Program implementers, recognizing that people who have left school prematurely are often difficult to reach, gradually shift their focus to serving current students with a high risk of dropping out. Any inability to focus earmarked resources on the target population is usually of great concern to external stakeholders.

3. *Does the number of clients served justify the program's existence?* Upon determining how many clients served come from the target population, the evaluator's next question elicits a judgment call about a program's performance, based on that number. Has the program served enough clients? This question is easy to answer *if* projected numbers were included in the program plan or other document. Some programs have a clear standard for client numbers over a given period of time. For example, a child care program might require that at least 20 children of mothers receiving welfare be enrolled in the daily-care program in any given week. In the absence of a clear criterion, the evaluator will need to consult with stakeholders about the issue before beginning the evaluation. Stakeholders usually can cite at least a general number considered an acceptable client base. Alternatively, the evaluator can use average numbers from comparable programs as a basis for determining whether the program is used enough. Insufficient numbers can lead to a declaration of the program's failure. For example, according to its program plan, a drug abuse prevention program might expect to pull in a few hundred youngsters with after-school, neighborhood-based activities. However, if just a few children turn out, it is not hard to view these numbers as a sign that the program has failed. The failure of a program is surmised if the program cannot reach a sufficient number of clients within the target population, cannot adequately screen them for appropriate intervention, or cannot retain most clients throughout the process.

An Example of Target Population Fidelity Evaluation. Among evaluators, a popular source for data about clients served is the records kept by implementing organizations. Glasgow, Lando, Hollis, McRae, and La Chance (1993) used client contact records to evaluate the reach of a smoker's hotline that provided a variety of smoking cessation services and was free to more than 2,100 health maintenance organization (HMO) members. The program was well promoted via a variety of channels, including newsletters and other mailings. Hotline scripts were pilot-tested, and experienced telephone staff and smoking cessation counselors were trained. During 33 months of operation, however, only 305 calls came in to the hotline, according to program records. That was an average of 2.3 calls per week (with 71% of all calls coming in just

after the program started), meaning the cost of each call to the hotline was an estimated \$81. The program was regarded as a failure because it was not reaching clients.

Fidelity Versus “Reinvention” in Conclusive Process Evaluation

The fidelity evaluation approach is predicated on the idea that the fidelity of implemented program components to intended components is highly desirable. It also assumes that less-than-complete fidelity reduces the effectiveness of programs in the field. These beliefs are not without controversy. There is within program evaluation a school of thought called the *diffusion tradition*, which argues that change is a necessary part of the adoption of any program (see Blakely et al., 1987). Such change is termed *reinvention*; its occurrence is, according to its proponents, absolutely necessary to preserve program effectiveness. Thus, discrepancy between a program plan and the observable implementation of the plan is *desirable* and should be encouraged.

Fidelity evaluation and the diffusion tradition are opposing viewpoints, but they can be somewhat reconciled by the typology and the contingency perspective as presented in this book. Reconciliation is accomplished in this way: In the initial implementation stage, a program needs frequent adjustment as it adapts to local circumstances; at this stage, therefore, reinvention is of benefit and can be facilitated by development-oriented process evaluation. Later, when a program is in the mature implementation stage or has operated in the field for a long period, making modifications can inhibit the smooth, efficient operation of that program. Also, implementation that deviates too much from the program plan can dilute the program's integrity. Thus, unless an undeniable reason for changing the program crops up during the mature implementation stage, the fidelity of the implementation to the intention should be protected. In the mature implementation stage, assessment-oriented process evaluation is appropriate, just as development-oriented process evaluation is appropriate in the earlier stage.

Given the reconciliation that the typology makes possible, it is not required that no change ever occur in the mature implementation stage. It is, however, assumed that stakeholders can justify any change that they make. If an HIV prevention program decided its target population would no longer be gay White men but instead would be gay African-American or Hispanic men, the change would have to be justified by data showing shifting demographics of the infection, or an evaluation showing that the program was more culturally

competent to serve members of minority groups, or some other compelling information. Funding agencies would not look upon such a change as being arbitrary, but documenting the change, and any others in the implementation, would be important for accountability purposes.

HYBRID PROCESS EVALUATION: THEORY-DRIVEN PROCESS EVALUATION

A conclusive process evaluation can provide information on whether an intervention has been implemented as intended. However, a hybrid process evaluation can also answer how the implementation of each element has contributed to the overall quality of the program. In other words, a hybrid process evaluation can serve both merit assessment and program improvement functions. This kind of information is very useful for understanding the general direction in which a program is moving and which program elements work or do not work well. More specifically, conclusive process evaluation, such as the fidelity evaluation approach found under the performance assessment strategy, typically takes on the assessment of just one or two action model components. However, in some situations, stakeholders may seek a complete assessment of the whole implementation of a program. Hybrid process evaluation based on the enlightenment assessment strategy as discussed in Chapter 2 fits this bill. Hybrid process evaluation can help determine the overall quality of implementation and fosters a systematic formulation of strategies—as opposed to piecemeal reactions—meant to improve the implementation.

The evaluation approach associated with hybrid process evaluation is called *theory-driven process evaluation* (see Table 2.1 for its place in the typology). Theory-driven process evaluation systematically assesses how the major components of a program plan are being implemented in the field. The technique can serve both program accountability and improvement functions. For example, if a program is found to have trouble retaining clients, theory-driven process evaluation can push the inquiry further to find out what is impeding retention. Theory-driven process evaluation basically uses program theory, especially its action model portion, as a framework for assessing the implementation process. Four issues are especially pressing when designing and conducting a theory-driven process evaluation. They range from communicating with stakeholders to combining qualitative and quantitative research methods, as explained below.

1. *Briefing stakeholders on the purposes, strategies, and procedures of evaluation.* Evaluators need to meet with key stakeholders to discuss the purposes,

strategies, and procedures of the upcoming evaluation. The meeting is a good opportunity to obtain stakeholders' support and hear their input.

2. *Clarifying the stakeholders' action model/change model schema concerning implementation.* Before proceeding with theory-driven process evaluation, stakeholders' action model/change model schema, especially as it pertains to the program plan and the action model, must be clearly communicated. (The material presented in Chapters 2 and 3 is a guide for evaluators beginning such a task with stakeholders.) For programs already implemented and essentially matured, agreement among stakeholders about what the program plan should look like usually comes fairly easily. But even if some components of the program plan do spark disagreement between key stakeholders, that is not an obstacle to evaluation. Rather, disagreement means that evaluators should test various hypotheses while investigating the implementation. Suppose key stakeholders in a program argue about who should be charged with implementing the program—professionals or trained peer volunteers? If implementers currently delivering services come from *both* these groups, evaluation can investigate the relative quality of service delivery by the two. Resulting data would be useful for settling differences among stakeholders as they continue planning future programs.

3. *Research methods for theory-driven process evaluation.* The action model/change model schema encompasses a variety of elements, and it often requires a combination of quantitative and qualitative methods of data collection.

4. *Freewheeling data collection.* The action model/change model schema is an effective guide as evaluators strive to focus on central issues. It is, again, a guide; it should not be regarded as a boundary, barring evaluators from examining important issues outside the framework. In fact, in the course of an investigation of the issues constituting the framework of the action model/change model schema, important questions beyond its scope tend to be generated (an illustration is included in Chapter 9). Pursuing these questions often provides further enlightenment information for stakeholders' use.

Examples of Theory-Driven Process Evaluation

Comprehensive theory-driven process evaluation is associated with certain strategies and approaches from the typology. Two evaluations are discussed here to illustrate some of the possible functions of this kind of evaluation.

Evaluating a Workplace Smoking Policy

Gottlieb, Lovato, Weinstein, Green, and Eriksen (1992) evaluated the implementation of a restrictive smoking policy for employees of a large, state-run human services agency. The program theory for this workplace policy included four elements: concept, context, process, and outcomes. Its terminology differs, but this model of the workplace smoking policy addresses issues similar to those in the conceptual framework that this book outlines. For example, the “concept” element consists of indicators such as goals/assumptions, nature of the policy and change, and development and support of the policy. These mirror to a degree the program rationale; intervention protocol; and, partially, the implementing organization, from the conceptual framework. Gottlieb and colleagues stated explicitly that their model was constructed based upon established literature; it is not clear, though, whether stakeholders’ input contributed to construction of the model. Gottlieb’s team used quantitative methods such as sampling and surveys to collect social and demographic information and review employee opinions on several issues: participation in policy development, compliance, and the policy’s impact. Qualitative methods such as group interviews, individual interviews, and solicitation of written comments were used as well to gather information about whether the policy had affected the relationship between smokers and nonsmokers, whether it had impacted smoking cessation, and how infractions had been managed. Findings for the four components follow:

- *Concept.* The program rationale was based on a needs assessment and was supported by public sentiment. However, a majority of employees said they had had little opportunity to provide input during the formulation of the policy.
- *Context.* The policy affected workers in offices, such as clerical staff, but did not strongly affect employees who spent most of the workday in the field. Furthermore, implementing the policy was easier for large work sites than for smaller ones.
- *Implementation process.* The great majority of respondents supported the policy, but many employees also said they were unwilling to report violations. Employees seriously doubted the confidentiality of their complaints. Many supervisors, too, were unwilling to report violations, particularly those by productive workers. The “designated smoking area” policy had generated some confusion. Issues raised included whether employees could move their work into a break room in order to smoke and whether nonsmokers should be compensated for unused smoking breaks.

- *Outcomes.* In general, nonsmokers perceived that air quality in their work areas had improved. Smokers, on the other hand, perceived that air quality in designated smoking areas had suffered. Nonsmokers' satisfaction with the policy had increased over time, whereas smokers' satisfaction had decreased.

Findings from the quantitative and qualitative data prompted the authors to suggest practical improvements to the policy and other similar ones. The suggestions were to (a) provide opportunities for employees to join in policy formulation and implementation, (b) provide training for middle managers on how to communicate the policy and enforce it, and (c) ensure that restrictions of smoking would be similar across job categories.

Evaluating an Anti-Drug Abuse Program

One comprehensive, theory-driven process evaluation that closely mirrors this book's conceptual framework of program theory is an evaluation of a large anti-drug abuse program for middle school students in Taiwan (Chen, 1997). The program asked teachers to identify drug-abusing students and provide them with counseling services. A small group of top officials within Taiwan's Ministry of Education had designed the program; under the nation's centralized education system, the Ministry of Education approved appointments and salaries of teachers and administrators. When the program began in January 1991, 3,850 students had been identified as active drug abusers. That number declined sharply, plunging 96%, to 154 students by June 1991.

The program's huge success led to a theory-driven process evaluation being conducted to examine how the program had been implemented. Hopes were that this program's example could be a model for the smooth implementation of other programs. The anti-drug abuse program featured a documented program plan, but it was incomplete in comparison to the action model or program plan illustrated in Table 7.1. Acting as facilitators, evaluators convened separate focus group meetings with top officials of the education ministry and with teacher representatives to obtain the information needed to complete the program plan. (The separate meetings acknowledged teachers' tendency to be silent in the presence of top officials, who have much more power than teachers do.) Evaluators played the role of facilitators and consultants, helping these key stakeholders develop their program theory. The final version of the program plan ultimately used for evaluation, presented on the left side of Table 7.1, was agreed to by both groups.

Table 7.1 The Spring Sun Program: Program Plan Versus Actual Implementation

<i>Program Components</i>	<i>Program Plan</i>	<i>Actual Implementation</i>
Target population	All drug-abusing students Drug use to be verified through urinalysis	Only those drug-abusing students who were easy to reach Urinalysis collection environment was not controlled
Implementers	Teachers provided with adequate drug abuse treatment training and information	Teachers lacked adequate drug abuse treatment training
Intervention protocol	Primary: High-quality counseling Secondary: Drug education classes	Counseling mainly involved use of admonishment, threats, and encouragement Drug education classes were offered
Service delivery protocol	Compulsory individual counseling	Compulsory individual counseling, but with problems such as lack of plan and objective
Implementing organization	Every school	Smaller schools had difficulty implementing the program
Linking with associate organizations	Effective centralized school system	Communication gap, mistrust between Ministry of Education and the schools
Ecological context: Micro	Eliminating video game arcades	Video game arcades still exist
Ecological context: Macro	Strong public support	Strong public support, but problematic education system (elitism)

SOURCE: Adapted from Chen (1997).

The program plan entailed mixing research methods—both quantitative and qualitative—to collect data. For example, quantitative methods were applied to rate teachers' satisfaction with a workshop on drug-counseling skills sponsored by the education ministry, whereas qualitative methods were used to probe contextual issues of the teachers' opinions of the workshop. The right side of

Table 7.1 displays empirical findings for the program's real-world implementation; comparison of the program theory to the implementation reveals large discrepancies. The program had been carried out, but the quality of services and the system of implementation were far from impressive. The discrepancies between plan and implementation resulted from a lack of appropriate counseling training, the overburdening of teachers with counseling work with no change to their usual teaching responsibilities, and a lack of communication as well as mistrust between an authoritarian ministry and the teachers. The evaluation results raised the question of how a program without strong implementation had achieved a 96% decrease in drug abuse in schools.

Evaluating an HIV-Prevention Intervention

There is increasing interest in using health care providers to deliver HIV-prevention services to their patients. Unfortunately, lack of counseling skills and time constraints within busy clinics serve as barriers to such efforts. The Providers Advocating for Sexual Health Initiative (PASHIN) intervention uses computer technology to assess each patient's risk behaviors and to determine his readiness for changing each behavior. Since the intervention does not require providers to spend time performing a detailed sexual-risk assessment and does not require providers to have received extensive counseling training, it has the potential to minimize some of the barriers associated with provider-delivered interventions. A theory-driven process evaluation was conducted to assess how the PASHIN intervention was implemented in the field (Chen et al., 2008).

The evaluators applied the action model/change model schema to evaluate the program. Since the evaluation focused on program implementation, evaluators applied the action model to guide the evaluation. Evaluators started by using the framework of the action model to facilitate key stakeholders in describing their program. Stakeholders identified the following components of the action model as relevant to their program:

Implementing Organization. The implementing organization is a university-based HIV/AIDS primary care clinic. The clinic has 11 attending physicians trained in infectious diseases. The primary care providers include nine infectious diseases fellows, two nurse practitioners, and one physician assistant. On average, the clinic sees approximately 80 male HIV-positive patients a week. In the past, not all providers considered HIV prevention to be a key focus of the clinic. The PASHIN project had strong support from top management and was the first major HIV-prevention effort by the clinic.

Associated Organizations/Partners. PASHIN was one of 15 sites participating in a demonstration project funded by the Health Resources and Services Administration. The coordinating center responsible for evaluating the cross-site project was Enhancing Prevention with Positives Evaluation Center (EPPEC).

Target Population. The target population of this program was HIV-positive men who have sex with other men (MSM) who received primary care at the university's HIV outpatient clinic. Thus, the risk assessment questions and the printed intervention messages were tailored to the MSM population.

Implementers. All providers (physicians) in the clinic agreed to participate in the initiative. They attended a 2-day training session to increase their HIV-prevention knowledge and communication skills. In addition, two additional training sessions were conducted. With regard to the implementers component, key stakeholders were particularly concerned with issues related to training, service delivery, and providers' beliefs about the effectiveness of behavioral intervention.

Based on the action model, key stakeholders and evaluators agreed to a set of questions that would guide the evaluation design and data collection:

1. *Did providers actually deliver the prevention session to their patients?* The key stakeholders wanted to determine the fidelity of the intervention, that is, whether the HIV-prevention interventions were actually delivered by providers to their patients. The evaluation method for answering this question was to ask each provider to write down the service delivery information immediately after seeing the patient, considering the following questions: Was an HIV-prevention session offered? If offered, what was the length of the session? Patients were asked the same questions in an exit survey conducted by a research assistant. The patient's service report provided a check on the accuracy of the provider's service report.

2. *What was the quality of the delivered HIV prevention sessions?* The quality of each HIV-prevention session was measured by asking a provider to rate the quality of the session after seeing a patient. The response categories of the quality scale were *excellent*, *good*, *fair*, *not so good*, and *terrible*. Patients were asked the same question in their exit survey.

3. *Did providers increase HIV-prevention behavioral counseling for all of their patients after the intervention was implemented?* Key stakeholders were interested in whether providers substantially increased their prevention activities with all of their patients after the intervention was implemented. The

measure of providers' HIV-prevention activities was a 4-item scale assessment developed by EPPEC. The assessment method used was the one-group pretest-posttest design.

4. *Were the trainings adequate?* The adequacy of training was measured by asking providers, using a Likert-type scale, to agree or disagree with the following statement: "I am confident that I have adequate training to provide HIV-prevention counseling to HIV-infected patients/clients." Providers were asked this question at baseline before the initial training and at the 6-, 12- and 18-month follow-up assessments.

5. *Did the intervention experience increase provider confidence in communicating HIV-prevention messages to patients?* Knowledge and skill deficits have been identified as major reasons providers may lack confidence in their ability to effectively deliver HIV-prevention interventions. Theoretically, the training and, especially, the experience of providing prevention counseling would increase providers' confidence in communicating about HIV prevention. Key stakeholders were interested in assessing this hypothesis: Providers will demonstrate increasing confidence over time in communicating HIV-prevention messages to patients the longer they deliver the PASHIN intervention.

6. *Did the intervention experience increase providers' favorable attitude toward the intervention or provider-delivered HIV-prevention interventions in general?* Key stakeholders theorized that the intervention experience would provide an opportunity for providers to appreciate the need for provider-delivered HIV behavioral prevention, which, in turn, would enhance favorable attitudes toward such an intervention approach. The evaluators and stakeholders were interested in assessing the following hypothesis: With increasing experience with the behavioral intervention, providers will demonstrate increasingly favorable beliefs about the effectiveness of the intervention over time.

Mixed methods were used to collect data to evaluate this component. The quantitative data were collected by surveying providers and patients at baseline and 6, 12 and 18 months. Qualitative data were collected by the following methods: (1) field diary (the project coordinators' day-to-day account of prevention activities in the clinic), (2) in-depth interviews with providers, and (3) in-depth interviews with the project coordinator. In accordance with the preference of interviewees, interviewers took notes rather than using a tape recorder during the interviews. The notes were coded based on manifest meanings, which were categorized to find themes in the dialogues.

7. *Did the intervention experience increase providers' belief in the effectiveness of behavioral interventions?* The research team theorized that experience with HIV prevention would lead providers to become aware of their clients' reduction in HIV-risk behaviors. As a result, the experience would enhance providers' belief in the effectiveness of behavioral interventions. Thus, the investigators assessed the following hypothesis: Providers will report more favorable beliefs in the effectiveness of behavioral intervention over time. Providers' belief was measured through their agreement (using the 7-item Likert-type scale mentioned above) with the following reverse-scored statement: "No matter how much you counsel some patients/clients with HIV, they are still going to infect others."

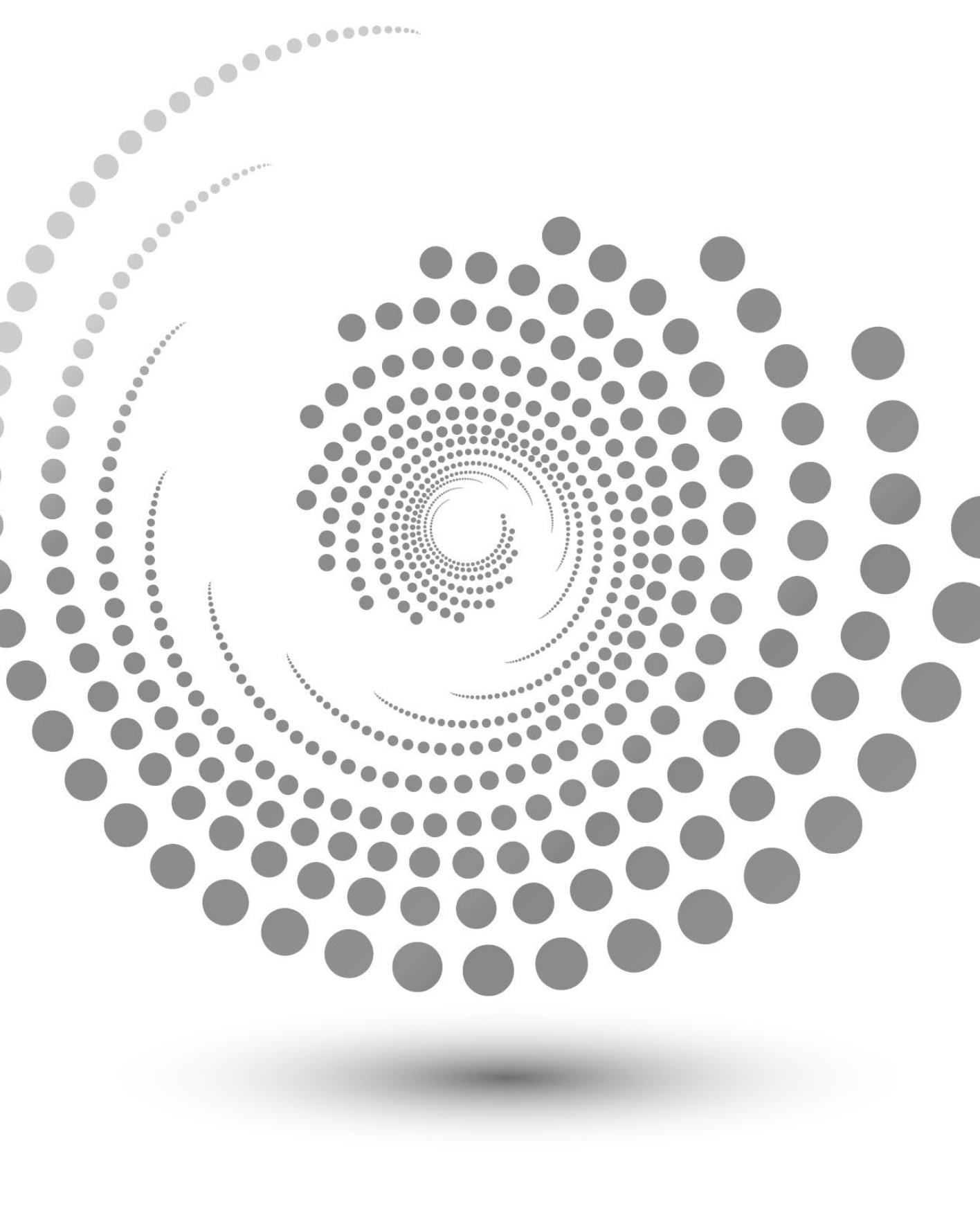
Overall, the results demonstrated that providers successfully delivered the computer-based, provider-delivered intervention within the context of regularly scheduled treatment sessions with HIV-positive MSM patients. The majority of providers (79.4%) and patients (83.5%) reported that the quality of HIV-prevention services delivered during these sessions was "good." The majority of the providers also reported that they had received adequate training, felt more confident in communicating about HIV-prevention issues with their patients, and provided more HIV-prevention counseling to their patients due to the project. However, the experience of delivering HIV-prevention counseling during an 18-month period did not appear to change providers' attitudes about provider-delivered HIV-prevention intervention or their belief in the effectiveness of HIV prevention in general. Patient counseling has always been an important component of the provider-patient interaction; however, reimbursement for the time spent on this activity is often not sought or may not be given if the proportion of time spent on counseling is not deemed critically important by third-party payers. Therefore, providers may not welcome the prospect of adding additional counseling to an already hectic clinic schedule.

Theory-Driven Process Evaluation and Unintended Effects

A significant advantage of theory-driven process evaluation is its capacity to detect positive and negative *unintended* effects, if these exist, due to its comprehensive examination of the implementation process. As a matter of fact, both the evaluation of the workplace smoking policy and the evaluation of the anti-drug abuse program discussed above turned up important, though unintended, effects.

QUESTIONS FOR REFLECTION

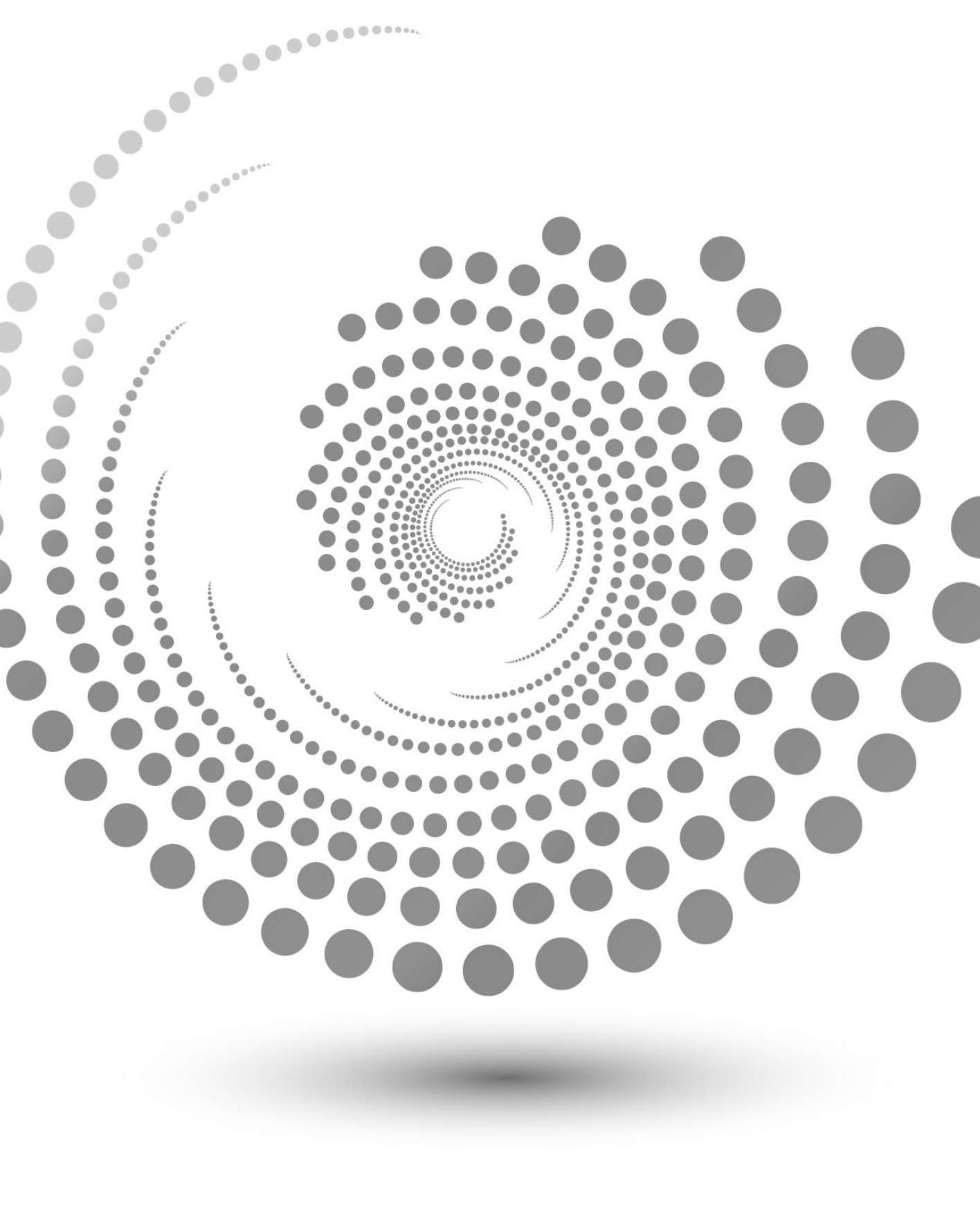
1. Discuss why constructive process evaluation may be needed in a mature program stage.
2. Compare and contrast the constructive process evaluation and the conclusive process evaluation. How are they similar? How do they differ?
3. Explain the different modes of accountability. Why must accountability incorporate these various modes?
4. Why is the fidelity evaluation approach valuable?
5. Discuss the four types of fidelity evaluation. Is it necessary to utilize all four types of fidelity evaluation when conducting an evaluation of a program? Explain.
6. Discuss the debate related to fidelity versus reinvention in the conclusive fidelity evaluation. Which side do you agree with? Why?
7. Discuss the conceptual framework used in theory-driven process evaluation? Explain why theory-driven evaluation is a hybrid evaluation.



PART IV

Program Monitoring and Outcome Evaluation

Here in Part IV, Chapter 8 contains a discussion of program monitoring, Chapter 9 offers a discussion of constructive outcome evaluation, and Chapters 10 and 11 address selecting an outcome evaluation strategy to gather credible evidence of a program's impact. Chapter 8 explains both process monitoring and outcome monitoring. Although, technically speaking, process monitoring properly belongs to Part III, because so many of the principles and techniques of process monitoring resemble those of outcome monitoring, it is convenient to discuss both in a single location.



CHAPTER 8



Program Monitoring and the Development of a Monitoring System

Stakeholders depend on some basic facts about a program's progress to inform their communication with funding agencies and other interested groups and for ongoing internal administrative purposes. Evaluators can help with the task of preparing to obtain these basic facts on a continual basis, a task called *program monitoring*. Once a program-monitoring routine has been developed, stakeholders themselves can perform the evaluation going forward. Program monitoring may well be the activity that best demonstrates the usefulness of empowerment evaluation. This is because in helping to launch ongoing program monitoring, evaluators both build the capacity of the program and its stakeholders to collect evaluative data and show them how to interpret and use their data correctly and meaningfully.

WHAT IS PROGRAM MONITORING?

An evaluation that calls for periodic collection of quantitative information about a program's process and outcomes is called *program monitoring*. Program monitoring is helpful as a provider of the kind of basic information to which nearly all programs wish to have access: a set of vital statistics concerning the program's implementation and outcomes. Data from program-monitoring evaluations are often used as indicators of program performance, but, really, program monitoring does not provide the in-depth information made available by other evaluation approaches, namely, process evaluation and outcome evaluation. Program monitoring is a useful counterpart to process evaluation and outcome evaluation, but it is not intended to replace them. The differences between monitoring

and evaluation will be discussed later in this chapter, which begins by introducing two common types of program monitoring: process monitoring and outcome monitoring. Process monitoring provides the basis for developing process evaluation, while outcome monitoring does the same for outcome evaluation.

PROCESS MONITORING

Process monitoring is the periodic collection of implementation information. However, there is a tremendous amount of information that can be gathered about implementation, too much for the evaluator to simply collect it all. Information about clients' characteristics and service delivery, for example, collected over a period of time, would rapidly become unwieldy. Therefore, evaluators need to understand what portion of the implementation information will best serve stakeholders' needs. The action model conceptual framework presented in Chapters 3 and 5 can help evaluators and stakeholders determine what information will be collected. In any case, process monitoring should, at a minimum, involve enough information to enable stakeholders and evaluators to know whether a program is serving the right individuals and whether they are receiving services as intended by the program designers. Process monitoring that gathers at least the following three kinds of information should prove useful.

1. *Sociodemographic backgrounds of clients.* An understanding of the variables of clients' backgrounds—their race, ethnicity, age, gender, education, and marital status—helps stakeholders see who is served by a program. The information may also detect any disparity in services.

2. *Kind of risk behavior clients represent and the severity of their need.* It is good for stakeholders to know as much as possible about the degree of risk behavior their targeted population of clients typically represents. For example, HIV-prevention programs meet with several categories of potential clients for whom interventions could be designed: men who have sex with men, intravenous drug users, commercial sex workers, and recent immigrants. In addition, the severity of various clients' needs is sure to differ, calling for various levels of intervention or treatment. The HIV-prevention program might find, for example, that clients in the commercial sex workers category rarely practice safe sex.

3. *Number of intervention activities clients have completed.* Stakeholders ordinarily like to keep abreast of information about clients' completion of the intervention activities. For interventions requiring multiple sessions, then, information such as the number of sessions completed by each client needs to be collected.

In this book, evaluators are urged to design the record-keeping system of all programs with which they work so that it supports process monitoring.

Uses of Process-Monitoring Data

Stakeholders such as funding agencies are given to asking the fundamental implementation questions: How many clients are served? Who are these clients? What kind of services are they receiving? For programs with process-monitoring systems in place, answering these questions is easy. Another advantage of process monitoring, due to the collection of standardized data at different points in time, is that the quality of implementation then and now can be compared: Was this program doing better at some point? Worse? A program that served 300 clients one year but only 200 the next may need some study by its stakeholders because the decline in client numbers could be an indication that something is wrong.

Process Monitoring Versus Process Evaluation

Process monitoring and process evaluation (e.g., fidelity evaluation and theory-driven process evaluation, discussed in Chapter 7) differ in important ways. Differences arise in three general areas: scope of data collection, depth of data collected, and data collection across time. The scope of data collection is much broader in process evaluation than in process monitoring. Process evaluation incorporates many aspects of the implementation process, such as the implementing organization, the ecological context, and the implementers. On the other hand, process monitoring seeks only *basic* information about client characteristics (sociodemographic data, relative risk) and client services (services provided, completion status). In addition, the depth of data collected is greater for process evaluation. Typically, there is little depth within process monitoring: Each person receiving a particular service is counted. No insight is generated into the meaning behind the numbers. If few minority clients were being served by a program, process monitoring would indicate their low numbers but say nothing about why they were low. By contrast, process evaluation would go further, providing in-depth analysis to obtain the contextual information capable of explaining the near absence of certain groups from the program.

Finally, data collection across time is a characteristic strength of process monitoring but *not* of process evaluation. Process monitoring collects data at various points across time, always with the same instruments that were used for the initial collection. This allows for comparison of a single program activity at one point in time with that same activity at another point in time. Comparative

information of this sort may not be generated by process evaluation unless its design incorporates process-monitoring techniques. In other words, process monitoring and process evaluation, used together, compensate for each other's weaknesses; many stakeholders will find this pairing worthwhile.

OUTCOME MONITORING

Outcome monitoring is the periodic collection of information about the outcomes of a program. The purpose of outcome monitoring is to acquire data to increase understanding of whether clients are better off for having received services (better off in terms of outcome measures). For example, outcome monitoring of an alcoholism treatment program would require repeated measures of clients' drinking behavior. Like process monitoring, outcome monitoring uses a standardized instrument to collect outcome data from clients at different points in time. For example, a program to reduce intolerance toward the HIV-positive population might develop an instrument that measures respondents' perceptions of and attitudes toward people infected with the virus. As the program conducts outcome monitoring, this standardized instrument would be used to measure respondents' levels of intolerance at least twice—once before an intervention and again after completion of that intervention. The instrument might also obtain clients' sociodemographic particulars.

Typically, outcome monitoring comprises four phases: identification of goals, identification of goal indicators and data sources, determination of needed background information, and pre- and postintervention collection of data.

Identification of Goals

To encourage stakeholder buy-in, the goals or outcomes to be scrutinized through the monitoring system are usually chosen by a committee of representatives of the various stakeholder groups. For example, a state's Tobacco Use Prevention and Control Task Force may charge a committee with setting overall goals for its statewide program. The committee may establish three goals:

1. Prevent youths (under age 19) from becoming users of tobacco products.
2. Promote treatment of tobacco dependency through providing increased access to cessation programs.
3. Reduce exposure to secondhand smoke.

Outcome Measures and Data Collection

To monitor goals, indicators and data for measuring them are needed. For example, say an antismoking task force found certain outcome measures available in existing data. To assess attainment of the goal of youths' avoidance of tobacco, the task force used a measure from the Youth Risk Behavior Survey (the reported percentage of youths who have never smoked tobacco). In many instances, however, data relating to outcome measures do not exist. For example, outcome data on participants in smoking-cessation programs is not widely available. For such an eventuality, evaluators should be prepared to devise new instruments for measuring outcomes and other data. A part of most outcome-monitoring projects is the collection of sociodemographic data about a program's clients. These data must be in hand so that the success rates of various subgroups within the clientele can be measured.

Outcome monitoring often requires a follow-up survey of clients. Collecting follow-up data on clients of programs that serve transient populations—drug abusers, sex workers, the homeless, and other high-risk groups—is especially difficult. Because members of the evaluation team must track and interview the program participants, the cost of the outcome monitoring contract is correspondingly higher. Thus, outcome monitoring is both more difficult and more expensive than process monitoring.

Outcome Monitoring Versus Outcome Evaluation

Stakeholders have a great interest in understanding the outcomes of a program. Outcome monitoring and outcome evaluation each provide useful information, but it is useful in different ways. As will be discussed in Chapters 10 and 11, the strength of outcome evaluation is its capability to provide evidence about whether an intervention is effective or not. Outcome monitoring does not, and this is the major weakness of outcome monitoring.

However, outcome monitoring has strengths. Although outcome monitoring is not meant to provide scientific evidence of program effectiveness, one should not underestimate its usefulness to stakeholders during program development. One of the strengths of outcome monitoring is that it provides data that respond, in a timely way, to stakeholders' concerns about clients' progress or lack of progress. Data of this nature will suffice when stakeholders want to strengthen a program, not sum up its effectiveness. Thus, if, according to outcomes, clients are not faring well following intervention, enough information exists to suggest that

the program is not working and that modifications that can strengthen the program should be identified. On the other hand, when outcome-monitoring data show that clients are experiencing better postintervention outcomes, stakeholders can justifiably conclude that the program is at least promising. The program may, in fact, have contributed to clients' progress, but an outcome evaluation will be necessary to answer categorically the question of program effectiveness.

Another strength of outcome monitoring that should not be overlooked is its affordability. Outcome evaluation is usually expensive and difficult. It is unreasonable to expect small, community-based organizations to treat their programs to outcome evaluations; the cost of conducting a randomized experiment can, in fact, equal some organizations' budgets for providing services. Therefore, funding agencies should be prepared to allow small organizations to rely on outcome monitoring rather than outcome evaluation. Furthermore, outcome monitoring functions as an evaluation capacity-building experience for organizations—a worthwhile investment. Indeed, outcome monitoring is a foundation for outcome evaluation. An organization that masters outcome monitoring is demonstrating a capacity to track client performance and use the data for program improvement. Once an agency can complete outcome monitoring, it will more easily appreciate and attempt outcome evaluation. Program staff who have experienced outcome monitoring will be less likely to resist an upgrade to outcome evaluation. Asking an organization to jump immediately into outcome evaluation, with all the associated disruptions, may be counterproductive if program managers and implementers believe the resulting data will not be very useful to their agendas.

Therefore, evaluators need to understand clearly the nature, functions, and limitations of the two activities. To summarize, a general rule for whether to conduct outcome monitoring or an outcome evaluation of an intervention program is this: When stakeholders need credible evidence of their program's effectiveness, choose outcome evaluation. When they want outcome data to show progress, to communicate with people interested in the program, or for internal management tasks, choose outcome monitoring. Outcome monitoring holds special relevance for organizations with tight budgets. In the same vein, a funding agency making a small grant to a community-based organization can reasonably request the agency to perform outcome monitoring, but expecting an outcome evaluation is unreasonable.

PROGRAM-MONITORING SYSTEMS WITHIN ORGANIZATIONS

Process monitoring is the gathering of data concerning the services clients receive from a program, whereas outcome monitoring is the gathering of data concerning

these same clients' outcomes following intervention. It makes sense to integrate these two in a *program-monitoring system*. Process monitoring and some outcome monitoring can be conducted simply with paper and pencil. However, an integrated program-monitoring system will not be so uncomplicated, usually requiring an electronic information system to store and manage data and build the organization's capacity to implement monitoring. This is the topic of the rest of this chapter.

With Congress's passage of the Government Performance and Results Act (GPRA) in 1993, there began a nationwide trend among federal, state, and local agencies to affirm more strongly their accountability for their programs. One impact of the GPRA is that more and more large agencies are instituting program monitoring, and even evaluation systems, to collect the data that assist them in meeting the requirements of the act. This development has also had a trickle-down effect among local agencies and community-based organizations. The program-monitoring systems being so widely adopted are *institutionalized* systems for gathering information, on an ongoing basis, about programs and their activities. In other words, the collection of monitoring data should become a routine task in an organization that implements a program-monitoring system. Monitoring data usually includes implementation and outcome data and can extend to the program-planning process. For example, to ensure that an organization's program-planning process is participatory, a program-monitoring system can observe whether, at the local level, community representatives and experts are being invited to join in planning.

Program-Monitoring System Elements

A program-monitoring system typically integrates four elements: monitoring guidance, an electronic data system, capacity building/technical assistance, and support from top management. *Monitoring guidance* is the process of outlining which standardized data elements will be collected, how they will be collected, how they will be reported to a central site(s), and how the information they generate will be utilized.

The *electronic data system* is required for most program-monitoring systems to enable them to process data. Given the current availability of information technology, it is feasible to build a web-based computerized system that electronically links all of an organization's local units to a central information center. Within such a system, data can be easily transmitted, stored, managed, and analyzed—at least when appropriate data managers are engaged.

This brings us to the element called *capacity building/technical assistance*. Maintaining a program-monitoring system depends on local units' and

implementers' mastery of the forms and other requirements mandated by whatever monitoring guidance is given. The local units and implementing staffs must integrate these requirements into their day-to-day activities. However, not every local unit and staff is immediately capable of doing so, and the first question asked must be whether the local organization has sufficient manpower to engage in such collaboration. A program-monitoring system requires designated staff responsible for caring for the system. It is not unusual for funds to be provided to local units that lack a staff to manage data. A second capacity-related question concerns whether the local staff has the skills and knowledge needed to collaborate. The quality of monitoring data stems from a staff's capacity and willingness to continually and correctly record their activities.

Most often, staff will need intensive training. Furthermore, because the first steps of launching a program-monitoring system within local facilities tend to be dogged by various difficulties, the larger organization should be ready with generous technical assistance for local staff. Such generosity is more forthcoming when creation of the program-monitoring system has the *support of top management*. Developing and launching a program-monitoring system—notably one at the state or national level—is a complicated process. It is likely to demand changes to the structure and activities of units across the entire organization. Predictably, such a top-to-bottom project will encounter problems, complaints, and resistance to change. With strong support from top management, these can be overcome; without it, the monitoring system threatens to fall apart after any stumble, and it can easily end up discontinued. Unless those with the greatest authority unquestionably back the program-monitoring system, an attempt to build it simply cannot be recommended.

Developing a Program-Monitoring System

There are two general strategies for developing a program-monitoring system. One strategy is to use a small group of experts and top managers to devise the system and ask employees down the ladder to implement it. An advantage of this approach is its efficiency. The experts need no more than several months to determine the instruments, data collection and reporting procedures, and staff training needed in order to be ready for implementation. This strategy has a disadvantage, however, in that implementers and administrators at the local level may not understand and support the system and thus commit to its launch.

In contrast, the participatory strategy consists of inviting representatives of various stakeholder groups to help develop the program-monitoring system.

Although the group nature of the endeavor entails more time and coordination of effort, when a system has been created in this way, implementers and other stakeholders are more likely to accept and support it. In addition, the presence of stakeholders' input tends to help the system respond better to local-level needs and interests, making it more useful to various organizational units. The system's capacity building/technical support also seems to thrive when the participatory approach is used. The participatory approach to creating a program-monitoring system is thus highly recommended. An example of this approach in action follows.

AN EXAMPLE OF DEVELOPING A PROGRAM MONITORING/EVALUATION SYSTEM

CDC-Funded Health Department HIV-Prevention Programs

In spite of a growing interest in building a monitoring system, little in the literature provides a concrete illustration of how to develop one. This section will provide a detailed discussion of the experience of developing a national monitoring/evaluation system to evaluate HIV-prevention programs funded by the Centers for Disease Control and Prevention (CDC; Chen, 2001; Glassman, Lacson, Collins, Hill, & Wan, 2002). This case study comes mainly from Chen's (2001) study.

Background

Since the 1980s, the CDC has funded 65 health departments to plan and implement HIV-prevention programs. In 1993, the CDC issued a supplemental guide to health departments, requiring them to formally include the local HIV-affected community in the planning of prevention efforts. Planning groups then worked with health departments to identify prevention interventions appropriate to the trends and needs of high-priority populations. The goal was to develop a comprehensive HIV-prevention plan that would address local situations with available resources. The plan, in turn, provides guidance to health departments for allocating resources and for working with community-based organizations to design and implement prevention programs.

Since 1993, the evaluation questions for HIV-prevention planning that includes the affected community have been evolving. In the first

(Continued)

(Continued)

few years, evaluation heavily emphasized issues related to implementing the requirements for community planning groups and the planning processes. Typical evaluation questions at that stage were, What is the composition of the planning groups? Are racial and ethnic minorities well represented in the groups? How are health department budgets being spent? Are the CDC funds being allocated at the local level according to priorities identified by the community planning groups? Does the planning process follow the community-planning guidelines established by the CDC? Common evaluation approaches and activities employed to answer these questions were case studies, to examine the community-planning process and group dynamics; self-assessment evaluation tools to empower states to conduct their own evaluations; and a survey across all states to assess expenditures and outcomes for the high-risk populations targeted.

However, as the community-inclusive HIV-prevention planning initiative has matured and become routine, the evaluation questions have expanded to include inquiries about program implementation and effectiveness. The following questions have been frequently raised: Do health departments actually implement the comprehensive HIV-prevention plans developed by community planning groups as they allocate resources and develop programming? What kinds of prevention services are provided? What is the quality of those services? Who receives those services? How can service delivery be strengthened? How can state health department staff learn from each other's implementation experiences? Do the prevention efforts reduce the incidence of risky behaviors, such as unprotected sex? Does the planning initiative or the programs that follow from it contribute to reduced HIV transmission? Federal, state, and local governments want to know the answers to these questions. HIV-prevention partners—such as health departments, community-based organizations, and the CDC—also want to know how they are doing and how they can strengthen the planning process and its resulting programs. Moreover, it is the CDC's responsibility to determine the status and progress of prevention efforts in all its funded health departments, not just in those health departments that have their own resources to develop and implement evaluation efforts.

These types of questions are difficult to answer without systematic, standardized approaches to documenting and assessing the implementation and effect of HIV-prevention efforts. To advance evaluation so as to improve programs and provide accountability at the local, state, and national levels, there was an urgent need to develop common strategies and measures that could be used to document and understand the varied and numerous programs health departments had funded and implemented. Beginning in the winter of 1997, the Program Evaluation Research Branch in the Division of HIV and AIDS Prevention (DHAP) at the CDC planned and developed a national system for evaluating a range of CDC-funded HIV-prevention activities on an ongoing basis.

Stakeholder Participation

To build such a monitoring system requires collaboration among and support from a large number of stakeholders. At least nine groups had to be involved. They included CDC administrators, CDC evaluators, CDC prevention program specialists, CDC capacity-building specialists, CDC data system specialists, health department administrators, health department evaluators or data management specialists, community-based organization administrators, and community-based organization evaluators or data management specialists. Because of the complexity of the task and the large number of interests involved, flexibility had to be a hallmark characteristic of everyone involved, including evaluators. A variety of communication modes, such as large formal meetings, conference calls, focus groups, small working groups, mail, and email, had to be used to varying degrees. In general, stakeholders, collaborating with evaluators, defined the purposes and scope of the evaluation, the evaluation questions to be addressed, the types of evaluation to be applied, the formats for collecting and reporting data, timelines for data submission, and how the evaluation mandates would be expressed in the funding announcements.

The stakeholders also needed to describe whether they had the evaluation capacity to meet the new evaluation requirements, how the CDC could help to build such capacity, and what kind of technical assistance they needed the CDC to provide. The CDC evaluators took the lead in technical

(Continued)

(Continued)

evaluation tasks, but they still had to collaborate continuously with local evaluators as well as other stakeholders on these issues.

Barriers to the Development of a National Monitoring/Evaluation System for HIV-Prevention Efforts

The development of the national evaluation system began by seeking consultations with key informants from the nine prevention partner groups described in the previous section. The informants agreed that there was an urgent need for the CDC to develop a national evaluation system. However, they also cautioned that the task would be highly complex and difficult. The major barriers they forecast are described below.

Barrier 1: Imposition of a Burden on Health Departments and Community-Based Organizations

The implementation of the new system would change the data collection activities and data-reporting systems of DHAP, the 65 CDC-funded health departments, and a few thousand community-based organizations (CBOs) that receive funds from health departments, as well as DHAP's own internal data system. Health departments and CBOs might have to revise their existing data systems or even develop new systems to accommodate the national system. They would also need to train staff to implement the system and ensure the quality of the evaluation data. Furthermore, many health departments would need to find ways to provide technical evaluation assistance to the local prevention providers they support.

Barrier 2: Fear of Arbitrary Use of Evaluation Results by the CDC

Health departments were worried that once the new evaluation system was set up, the CDC would use the findings as punitive criteria when awarding grants. Those departments performing poorly, as determined through analyses of these newly available data, would, they feared, receive substantially reduced funds from the CDC, without regard to the complex factors and situations that affected HIV prevention in a local area and regardless of how justifiable their level of performance was. Those states that had less funding than others because of a low prevalence of AIDS in their jurisdiction were particularly worried about this issue.

Barrier 3: Concern About Low Input From Stakeholders in the Development of Guidance

Health departments and other stakeholders were concerned that the CDC would take a top-down approach to developing guidance. In other words, they believed that the CDC might develop guidance without consulting them and simply mandate its implementation. They were troubled that, despite the CDC's best efforts, the evaluation guidance developed through such a top-down approach would be difficult to implement, provide misleading information, and generally be useless to them.

Barrier 4: Lack of Expertise and Capacity in Health Departments to Implement the New Requirements

An evaluation system such as the one proposed would "raise the bar" in terms of expectations for evaluation. There would be a concomitant need for health departments to have evaluators and electronic information experts to implement the system. Health departments were concerned that they would not have the financial resources, staff, or expertise to implement the guidance, and they did not know how to obtain these resources. The health departments thus believed that the requirement to implement the evaluation system would set them up for failure.

Barrier 5: Concerns Within the CDC

There was skepticism within the CDC regarding the feasibility of engaging in such a big, complex task. There was no precedent to follow and no guide for developing such an evaluation system for HIV prevention. To complicate matters, the branch responsible for developing guidance and the overall system (i.e., the Program Evaluation Research Branch) was a new branch and had not yet established itself within the CDC.

Building Momentum for Collaboration and Surmounting Barriers

Skepticism on the part of the AIDS directors, who supervised HIV-prevention and care-related activities in health departments, and their staff did not connote conceptual rejection of the development of a national evaluation system. They agreed with the CDC that such a system would benefit their

(Continued)

(Continued)

programs and overall HIV-prevention efforts. They were concerned, however, that the CDC was developing a rough draft without full and formal collaboration with them. The CDC's efforts to assure them that there would be full collaboration once a rough draft was available for review was viewed as disingenuous (as discussed later). Health department directors and staff believed that they were responsible for service delivery, knew the evaluation needs, understood how to change existing information systems to meet the new requirements, and had the necessary firsthand knowledge of their individual strengths and limitations to collect evaluation data. Further, they believed that without their assistance in its development, the system would become a data collection exercise disassociated from the needs of programs. Fortunately, their strong desire to produce an evaluation system that would be practical and useful in fact coincided with the CDC's intention and goal.

The organization most responsible for facilitating the transition from common goals to active collaboration was the National Alliance of State and Territorial AIDS Directors (NASTAD), which represents the health department AIDS directors. After the rough draft was finished, NASTAD worked with the CDC to convene several meetings where the AIDS directors and selected health department staff met with CDC staff to discuss the guidance. In these meetings, issues such as purpose and content, the required data elements, methodologies, feasibility, and technical assistance were discussed. The NASTAD facilitated discussions about the data elements that should be collected to adequately address the evaluation questions and the feasibility of collecting these elements. Revisions to the guidance were made after each meeting, and each revision was sent to all health departments for comment.

In addition, the CDC formally consulted with other stakeholder groups. This included an expert panel of evaluators who commented on the soundness of the evaluation logic and methodology. Because the health departments fund CBOs to deliver prevention services and would need data from them to meet the reporting stipulations of the system, a panel of CBO representatives was also convened to make suggestions.

Including stakeholders in the development of evaluation guidance alleviated the concern that the CDC would take a purely top-down approach to developing guidance (barrier 3). These collaboration meetings also provided a vehicle for stakeholders and the CDC to discuss other barriers and

build consensus on how to surmount them. For example, the concern about the burden on health departments and their CBOs (barrier 1) was ameliorated by the collective recognition that the CDC simply could not meet its accountability requirements without a national monitoring/evaluation system. Failure of the CDC to meet these requirements could lead to disaffection among its supporters and might even endanger the overall resource commitment of the nation to HIV prevention. All recognized the need to compromise and to move forward to build the capacity to conduct evaluation throughout the national, state, and local programs.

The concern about the potential punitive usage of evaluation data (barrier 2) was alleviated by the mutual understandings between the CDC and stakeholders that developed during the meetings and subsequent training workshops for health department staff. A mutual understanding developed that program effectiveness is determined by many factors (such as the state-of-the-art intervention technology and the shifting of epidemic patterns) that are not under the control of health departments and CBOs; all parties understood that it would be inappropriate to use evaluation findings as the sole criterion for determining or reducing funding. Further, a shared understanding developed that these evaluation activities were an important activity for program improvement. The discussions regarding the lack of capacity within health departments to conduct evaluation (barrier 4) led the CDC to speed up preparation and implementation of capacity building and technical assistance activities to support evaluation guidance. During these meetings, many AIDS directors and health department staff shifted from being skeptical to understanding the rationale of such guidance and the proposed evaluation system. Participants provided a wealth of suggestions and assistance in developing guidance and in testing the system.

Meanwhile, intensive cross-branch efforts within DHAP built support for the enterprise and encouraged collaboration in the development of evaluation guidance and the data system. Ongoing communication and joint efforts across DHAP branches were important to alleviating the concern that the undertaking was beyond the ability of the evaluation branch (barrier 5). Collaboration with three branches was particularly crucial. The first was the Prevention Program Branch, which is responsible for providing and overseeing the CDC HIV-prevention funding allocated to health departments. The implementation of the guidance would need the Prevention Program

(Continued)

(Continued)

Branch's cooperation and support for including evaluation system requirements in the funding announcement. Staff from this branch would also be the first line of contact for health departments regarding their evaluation technical assistance needs. The second crucial branch was the Capacity Building Branch. To implement guidance, the CDC needed to train technical assistance providers to help implement its requirements; this task fell to the Capacity Building Branch. The third crucial branch was the Statistics and Data Management Branch. This branch facilitates and oversees the quality and consistency of DHAP data systems.

Many cross-branch meetings were held, and the participation of these branches in the early stages of the development process bolstered their support for the system. Branch chiefs and staff provided support and useful advice and assistance in developing the guidance. The development of both the guidance and the data system also benefited from the consistent support of the division director.

Principles for Developing Evaluation Guidance

During the meetings and communications described above, several important principles were identified to guide future development of evaluation guidance and the overall system.

Make Guidance Useful for Both Program Accountability and Improvement

The guidance and its implementation would require the CDC and health departments to commit considerable resources and effort. Participants at the meetings strongly asserted that guidance had to be useful in meeting health department program improvement needs as well as the CDC's need for accountability. To be beneficial, the guidance needed to systematically assess the crucial components of programs across planning, implementation, and outcome.

Satisfy the Need to Aggregate Data at the National Level

A major purpose of the evaluation system was to create a data system that would provide the CDC with the national-level data it needed to produce reports for government decision makers. The system would account for

the effects that expenditure of national resources on HIV prevention had on the determinants of transmission. The health departments would have to collect standardized data to facilitate aggregation, analysis, and comparison across reporting units, thereby generating national-level data. Thus, the emphasis of the guidance had to be on collecting quantitative rather than qualitative data. The guidance was intended to describe the minimum data required to meet basic accountability and program improvement needs, not to be the gold standard for all possible evaluation activities. The guidance recognized the usefulness of qualitative data in evaluation at the local level but acknowledged that providing comprehensive qualitative data at the national level is beyond the capacity of the evaluation system. Health departments were still to conduct their own qualitative data collections and evaluation as needed. For example, a health department might need to conduct a case study to collect detailed information on how its outreach programs were implemented.

Pilot-Test the Guidance

The guidance was to be tested in several health departments to determine whether it was practical and feasible. The pilot test would also give the health departments experience they could use to inform suggestions for fine-tuning the guidance.

Format the Guidance to Increase Acceptability

The guidance was originally provided in a large volume containing the data requirements and information for submitting the data to the CDC, as well as detailed instructions on and procedures for collecting the data. The participants in the meetings said that it was difficult for them to use this huge document. They preferred to have a concise volume that clearly specified what the health department was expected to deliver to the CDC. The supplemental materials on how to collect the data, according to the participants, would be best put in a separate volume.

Phase In the Implementation

Many types of evaluation were addressed in the guidance volume. Some would take more time to prepare for than others. Because health

(Continued)

(Continued)

departments varied greatly in their capacity to collect and use evaluation data, a phased approach to implementing the guidance was agreed upon. Health departments would submit evaluation information on the planning stage first, followed by process evaluation information and, finally, outcome evaluation information.

Determine Required Versus Optional Evaluations

The original draft of the guidance volume stipulated various types of evaluation for assessing planning, implementation, and effectiveness. Stakeholders were concerned that they would not be capable of implementing all of these different types of evaluation during the 5-year funding cycle, even if a phased approach was used. After considerable discussion, it was agreed that two types of evaluation would be made optional. One was a survey of community planning groups regarding their opinion of how well the planning process worked. Health departments pointed out that many of them were already conducting such assessments and it would be difficult to standardize this largely qualitative information. Moreover, the survey described in the guidance volume would not provide health departments with enough additional information to be worth the effort.

The other type of evaluation made optional was outcome monitoring. Outcome monitoring in this document referred to tracking the progress of clients or a program based upon outcome measures set forth in program goals, without isolating the causes of this progress. As stated, outcome monitoring is different from outcome evaluation. Generally speaking, outcome evaluation entails the application of rigorous methods, such as experimental and quasi-experimental designs, to provide credible data on whether a program has attained a predetermined set of goals. Although all parties agreed that tracking risky behaviors and other determinants of HIV transmission would be valuable, the health departments did not believe they had the capacity to implement widespread monitoring of the outcomes of interventions during this round of funding.

Provide Technical Assistance and Capacity Building

Health departments anticipated needing considerable technical and resource support from the CDC to develop and implement their evaluation systems. Stakeholders strongly recommended that the CDC develop a

coherent technical assistance system to provide expertise and training and that the CDC supply additional funds to build health department capacity to implement the guidance.

The Content of the Evaluation Guidance Volume

Five states and one city volunteered to participate in pilot testing the guidance volume. The results of the pilot test, along with the principles described in the previous section, were used in the development of the final version of the guidance volume. The evaluation guidance volume was released in two parts: (1) guidance and (2) supplemental materials. The latter was referred to as a "supplemental handbook." For each chapter in the guidance volume, there was a corresponding chapter in the supplemental handbook. The guidance volume addressed the following topics: evaluating community-inclusive HIV-prevention planning; designing and evaluating intervention plans; monitoring and evaluating the implementation of HIV-prevention programs; evaluating linkages among the comprehensive HIV-prevention plans, the CDC funding application, and resource allocation; evaluating outcomes of HIV-prevention programs; and developing an evaluation plan. A phased-in approach was used to implement the activities described in the guidance volume. Data on designing and evaluating intervention plans and evaluating linkages between plans and resource allocation were submitted in September 2000. The implementation data was to be submitted in September 2001 and outcome evaluation data in September 2002.

Technical Assistance and Capacity Building

The final draft of the evaluation documents was finished in December 1999. Reaching this stage took about 2 years. Right after the completion of the final draft, the following strategies and activities were used to meet health departments' needs for technical assistance and to improve their capacity to conduct evaluation.

Training of Health Department Staff

Training sessions were offered to health department staff. The primary purpose of these training sessions was to assist health departments in implementing the evaluation guidance by increasing staff familiarity with the

(Continued)

(Continued)

contents of the documents. Secondary goals were to facilitate information and materials sharing, to enhance networking among health department staff and evaluators, and to identify specific technical assistance needs. The participants selected to attend were those who would be operationally responsible for implementing the guidance. They were assigned to a specific training session based upon an assessment of the evaluation capacity of the health department in which they worked. Capacity was stratified through an assessment of factors such as resources earmarked for evaluation, staff skill, and experience.

Peer Resource Manual

A few of the higher-capacity health departments implemented the evaluation guidance even before the training sessions. They were asked to document this experience. With the assistance of a contractor, those experiences were to be compiled into a "lessons learned" document.

Technical Assistance System

Within the CDC, there were intensive meetings among branches regarding the establishment of a technical assistance system to help health departments implement the evaluation guidance. The tentative scheme was that a request from a grantee health department would go directly to the CDC project officer for that health department. The project officer would refer the request to a technical assistance team in the Capacity Building Branch, which in turn would coordinate with other branches or external organizations to provide needed services. A tracking system was planned to monitor and ensure the quality of the technical assistance provided.

Increased Funds to Strengthen Evaluation Capacity

To enhance the capacity of health departments to conduct evaluation activities, the CDC provided new funds specifically designated for evaluation activities. These supplements included a base award to all health departments of \$50,000 and an additional \$100,000 to those who did not receive the 1997 evaluation supplement and that currently receive a total HIV prevention award of more than \$1 million from the CDC.

Increased CDC Data Capacity

The Statistics and Data Management Branch began to develop a data system to which health departments could submit data through the Internet. Also, the capacity of the Program Evaluation Research Branch to collect, store, manage, analyze, and report evaluation data was augmented with new staff and improved information systems.

Issues With Stakeholder Participation

The experience of developing this evaluation system may provide insights into issues related to stakeholder participation in developing a large evaluation system.

Timing of Stakeholders' Involvement

The participatory approach to evaluation requires ongoing interactions between evaluators and key stakeholders. In developing a large-scale evaluation system, several formal meetings are necessary to fully discuss the more complex and contentious issues. Meetings of this nature usually require a long time to plan and schedule. Along with formal meetings, the participatory process also requires regular and frequent audio conferences and communications through the mail, email, fax, and telephone. A draft needs to be revised many times until it is satisfactory to stakeholders and evaluators. It is undeniable that stakeholder participation considerably slows down the process. On the other hand, stakeholder participation turns a potential academic exercise into a process that produces a product that people are willing to support and use.

One of the more challenging tasks in developing a large system is to balance participation with a sense of urgency in such a way that stakeholders will acknowledge the need to complete the project in a timely manner. In planning for the development of the evaluation guidance, evaluators were torn between these competing demands. Two options were contemplated by CDC evaluators, each with its own strengths and limitations.

Option 1: Stakeholders Participate From the Outset

Stakeholders, such as AIDS directors and state health department staff, had been advocating for some time to be consulted during predesign stages;

(Continued)

(Continued)

they sought to be asked whether a project should even be considered. They could have been invited to participate in every stage of the development of the evaluation system, from conceptualization to the end product. The merit of this option is that stakeholders feel consulted from the onset and consider themselves full participants. This might have swiftly removed many of the barriers discussed above. However, state health departments are highly divergent in their perceptions, interests, expertise, and capability to conduct program evaluation. Involving them before there was a written draft to focus the discussion could have been extremely time-consuming and difficult. Many meetings over an extended period might have been required simply to discuss the development process. Similarly, abstract discussions of the potential content, methodology, instruments, and data sets could also have become protracted or even reached an impasse.

Option 2: Stakeholders Participate After CDC Evaluators Develop a Draft

With the second option, the CDC evaluation team could develop an initial draft based upon their best knowledge of evaluation theory, HIV-prevention efforts at the state and local levels, CDC and stakeholder needs, and some informal consultations with stakeholders. Stakeholders would then formally be brought in to comment on the draft. At that point, the evaluators would work closely with the stakeholders to revise the guidance volume until an agreed-upon version had been developed. The merit of this approach is that finishing the project would take less time. Rather than discussing abstract ideas from scratch, stakeholders and evaluators would have a concrete draft upon which to build. Suggestions would be specific, and changes would be easier to make. The disadvantage of this approach is that stakeholders would feel they were not fully consulted, especially in the early stages. Even if they were informed of the strategy for developing the guidance volume, they might be suspicious of its "real" motive and purpose, and those suspicions could raise concerns regarding development of the guidance at all.

After serious consideration of the pros and cons of both options, in the end the decision was to pursue the second. The rationale underpinning that decision was the goal to have the guidance finalized and integrated into CDC-funded state/local HIV-prevention programs at the inception of an upcoming 5-year funding cycle.

Stakeholders reacted to this course of action with uneasiness. Efforts by the CDC team to explain the strategy and ask for time to develop a draft met with much criticism from health departments and other stakeholders. However, things went better once the consultation process was underway, as comments made by stakeholders in the subsequent meetings and via other modes of communications were incorporated. Stakeholders actively participated in revising the draft. Many helpful suggestions on how to build a data system around the guidance volume were offered. In the end, the draft did help stakeholders to provide systematic and concrete comments.

However, some AIDS directors have a slightly different account of the evolution of the draft. They feel that the lack of collaboration in the very early stages led to weaknesses in the original draft, and they believe that their input greatly enhanced the quality and feasibility of the guidance volume. Nevertheless, there is no dispute that, in the end, both health departments and the CDC were pleased with the ultimate product. One lesson the CDC staff learned from this experience is that there were not simply two options for the development of the guidance but rather a continuum of options.

While there were pressing time constraints, the long waiting period between when stakeholders first learned of the decision to develop the guidance volume and the completion of a draft for formal collaboration deepened skepticism and generated much criticism. In hindsight, the CDC team should have requested that stakeholders organize a small working group to represent them in the early drafting process. The CDC staff could then have had telephone conferences with the working group periodically to seek its advice and to report the progress of the draft. The working group would have lent credibility to the CDC's assurance that stakeholder concerns and perspectives were being considered and that there would be opportunity for everyone to comment before anything was finished. This alternative approach would have balanced the desire for broad involvement of stakeholders and for finishing the work in a time consistent with programmatic needs, that is, the onset of a new funding cycle.

Effect of Stakeholder Participation

The CDC and health departments had communicated continuously about issues related to the development of the evaluation system through

(Continued)

(Continued)

different channels, such as meetings, working groups, conference calls, and so forth. The representatives of the health departments who participated in this communication were usually selected by NASTAD. These communications provided ample opportunities for representatives to express their perspective on what did or did not work well. The CDC took the feedback into consideration when planning future actions. In general, the feedback showed that, as a result of the participatory process, stakeholders supported the evaluation, participated in decision making, and provided useful input throughout the developmental process of the national system.

However, stakeholder participation also means that give-and-take occurs. Evaluators cannot get everything they want. It is important to understand the effect of including negotiation in the process. The experience of developing this evaluation system sheds some light on this issue. CDC evaluators made compromises related to issues of scope and focus, such as which parts of the evaluation could be done now and which parts would have to wait, which evaluation questions needed to be rephrased to reflect stakeholders' reality, what kind of assistance health departments would need, and how evaluation data should be used. Specifically, as mentioned earlier, evaluators gave up the survey of community planning groups and outcome monitoring. On the other hand, there was little dispute about issues related to the integrity of the design and methodology, which were of most concern to evaluators. For example, stakeholders had no objection to the evaluators' proposal that rigorous quantitative designs, such as experimental and quasi-experimental designs, be used to evaluate the outcomes of their programs. Stakeholders also wanted credible data to inform them of what worked and did not work so that they could strengthen their programs.

It is not clear how generalizable this experience is. For future development of the stakeholder participation approach, more information about diverse experiences and further study of the issues involved are needed.

Lessons Learned

Several important lessons were learned from this experience that may be useful to any organization contemplating or developing a large monitoring/evaluation system.

Development of a Large Evaluation System Requires Stakeholder Participation

A large evaluation system requires stakeholders to provide data. In the existing political structure, an evaluation system that is developed exclusively by a federal agency for states to implement would not work. However, the experience related in this article clearly demonstrates that a stakeholder participation approach can result in stakeholder buy-in and support. In some situations, an organization technically has the authority to build and implement a large evaluation system without stakeholder participation. However, with so much skepticism and even fear in the minds of so many stakeholders, such an approach might convince stakeholders to provide invalid data to the system to protect themselves. As a result, the evaluation system developed would be neither realistic nor useful.

Strong Commitment Within the Organization Is Required

The advent of a large evaluation system necessitates substantial changes in structures and activities throughout the affected organizations. In the case of this evaluation, substantial changes were made in the CDC, health departments, and community-based organizations. With change of such magnitude, a project predictably will run into problems, and stakeholders may complain, resist the changes, or be unable to meet the challenges. Unless strong leadership is in place to push the project forward, and unless top management within the affected organizations delivers unstinting support and commitment, the task can easily stumble and succumb to difficulties before its fruition.

Capacity Building and Technical Support Should Be Part of the System

It is the stakeholders who will implement the system. Stakeholder support for the evaluation system depends, in part, on whether resources and assistance will be provided to build their capacity and help them to deal with problems. Therefore, an evaluation system requires the systematic integration of guidance, capacity building, and assistance. The experience of developing this evaluation system indicates that stakeholders can be

(Continued)

(Continued)

insightful and helpful in bringing about such integration. As of this writing, although a portion of the data specified in the guidance has been submitted to the CDC, the generation of useful information from the new evaluation system for serving program accountability and improvement needs is still some time off.

However, noticeable structural changes in health departments and at the CDC are attributable to the guidance and the process of its development. Immediately, health departments began to develop or upgrade their information systems and to link them with CBO systems, to network with the local evaluation community, and to train health department and CBO staff on program evaluation as well as on the implementation of the guidance. Many evaluators were consulted with or hired to develop and implement the health department evaluation systems. In addition, several states reported during the annual HIV-prevention conference of March 2001 that they had already put their evaluation systems together and had begun to implement the evaluation guidance ahead of the requirement to do so. In general, they reported that building their new evaluation system had been laborious and challenging, but rewarding. The data collected up to this point have been used intensively by community planning groups in developing their comprehensive HIV prevention plans. States also indicated that, because of the guidance, the quality of intervention plans submitted this year by CBOs was much better than in previous years. A formal and systematic evaluation of the effects of the evaluation guidance was initiated by the Program Evaluation Research Branch at CDC.

Within the CDC, changes attributable to the evaluation guidance also appeared. Because of the experience of developing this evaluation guidance and system, DHAP wants to expand evaluation activities to other prevention service providers. Staff members from evaluation, program, capacity-building, and data management branches are currently working together to develop evaluation guidance and data systems for the community-based organizations that deliver prevention services, as well as for the national and regional organizations that provide technical assistance and capacity-building services to those community-based organizations.

QUESTIONS FOR REFLECTION

1. What does the process of program monitoring allow stakeholders to do? How can evaluators help with this process? What information is assessed with this technique?
2. Compare and contrast process monitoring and process evaluation. Give examples of each.
3. Can all information gathered during process monitoring be used in outcome monitoring? Why or why not?
4. Why is it important to measure goals? What tools can evaluators use in this endeavor?
5. What are the strengths and weaknesses of outcome monitoring?
6. List the similarities and differences between outcome monitoring and outcome evaluation.
7. Why is it inappropriate to measure the effectiveness of an intervention using outcome monitoring?
8. Why is outcome monitoring a relatively affordable activity?
9. List examples of real-life program-monitoring programs used within organizations.
10. Research the Government Performance and Results Act (GPRA). Why was it important for the government to enact legislation regarding program accountability?
11. Discuss the four integrated elements that make up a program-monitoring system.
12. Discuss the major barriers to building a monitoring/evaluation system in the case study of HIV prevention. Discuss also the strategies used to address these barriers.
13. Why is it important to have stakeholder involvement in the development of an evaluation system, even if it may considerably slow down the development process?

CHAPTER 9



Constructive Outcome Evaluations

As discussed in Chapter 1, there are two types of outcome evaluations: constructive and conclusive. The purpose of constructive outcome evaluation is to provide information for improving program outcomes. Conclusive outcome evaluation has a different purpose: It provides a formal assessment to determine whether a program has a desirable effect on its stated goals or outcomes. This chapter focuses on constructive outcome evaluation. Three approaches to constructive outcome evaluation will be introduced: SMART goals, the evaluability assessment, and the plausibility assessment/consensus-building approach. A brief preview of conclusive outcome evaluation will be provided at the end of the chapter. The purpose of the preview is to give evaluators and stakeholders an introduction to two options for designing and conducting a conclusive outcome evaluation. They can select the appropriate one to use based on their needs. Chapters 10 and 11 will then provide a detailed discussion of the experimentation evaluation and real-world effectuality evaluation approaches.

CONSTRUCTIVE OUTCOME EVALUATION

Constructive outcome evaluation can take many forms. A simple example of constructive outcome evaluation is when managers of a large rental property ask evaluators to survey tenants about their satisfaction with factors such as cleanness in the common areas, security, trash collection, noise, responses to requests for service, landscaping maintenance,

playground facilities, and so on. The information will be used to improve the rental property, which will in turn enhance tenants' satisfaction.

However, measuring an intervention program's outcomes is not so straightforward. A program may not have a set of goals or outcomes that are appropriate for evaluation purposes. Sometimes, a program may not have clear, measurable goals or outcomes. For example, an elderly service program might have a goal of enhancing senior citizens' social life but lack a definition of "quality of social life" or a way to measure it. Alternatively, management and staff may not share the same goals or outcomes. For example, managers of a welfare-to-work program may want counselors to spend more time and effort on more difficult cases, while counselors may want to spend more time with cases that are more likely to have a positive outcome.

In these cases, constructive outcome evaluation is used to facilitate stakeholders in developing a set of appropriate goals, to strengthen the coherence of the program's structure, to solidify consensus on a set of goals for evaluation purposes, or to prepare a program for conclusive outcome evaluation. As will be discussed later, the literature on evaluability assessment (Smith, 1989; Wholey, 1987, 1994) has made an important contribution to this kind of assessment. However, there are other tools that are useful to evaluators as well. This section will introduce three evaluation approaches under the umbrella of constructive outcome evaluation: SMART goals, evaluability assessment, and plausibility inquiry/consensus building. These three tools are related but have different emphases, as follows:

- *SMART goals.* This tool is particularly useful for developing specific and measurable goals.
- *Evaluability assessment.* This tool is an intensive inquiry into whether a program has measurable goals and whether resources, management, and implementation are organized in a way that facilitates goal attainment.
- *Plausibility assessment/consensus-building approach.* This tool is useful for identifying plausible goals and outcomes for evaluation and building stakeholder consensus about them. This tool is particularly useful for programs of large scale or with broad aims.

SMART GOALS

SMART stands for the requirements that goals be Specific, Measurable, Attainable, Relevant, and Time-bound. SMART goals, first introduced by

Doran (1981), have become a popular tool. Management uses SMART goals as a guide to developing realistic and measurable goals or objectives. SMART goals are also useful for social betterment and health promotion programs, which usually have general goals that describe what the program wants to accomplish. Evaluators may use SMART goals to assist stakeholders with developing a set of measurable goals for an intervention program.

Following is a description of each characteristic of SMART goals.

Specific

The goal needs to be clear. A goal is specific when it can at a minimum answer these questions: “What exactly do we want to accomplish?” and “For whom exactly?” For example, a statement such as “provide services to needy people” is not specific because it provides little information about what “services” will be provided. It also provides no information about who the “needy people” are. On the other hand, a statement such as “provide counseling and shelter assistance to victims of family violence” is specific because it unambiguously states what will be done and for whom.

Measurable

Progress toward a goal must be measurable, and measurement data must be collected. For example, a statement such as “improving client satisfaction” is less measurable than a statement like “reduce client waiting time from 10 minutes to 3 minutes.”

Attainable

A goal must reasonably be attainable with existing resources. A goal can be stretched slightly so that a program’s staff feels challenged, but the challenge must not be too extreme. Setting an impractical goal demoralizes staff and makes the goal meaningless. For example, if the chair of a university’s social science department asked faculty members to have four articles published each year but did not provide assistance toward that goal, such as a reduced teaching load or more funding for graduate assistants, that goal would be unrealistic.

Relevant

A goal must relate to the overall purpose of a program. For example, if a professional development meeting for government workers had as a goal

“spiritual enhancement,” that goal would justifiably be criticized as irrelevant to the meeting’s purpose of developing professional knowledge and skill.

Time-Bound

Goals should be linked to a specific time frame to promote a sense of urgency. If a goal is not time-bound, the goal is unlikely to be achieved. For example, a goal statement such as “conduct a workshop” is not time-bound as is the goal statement “conduct a workshop by September 2014.” The latter statement is more effective than the former in mobilizing staff to achieve the goal.

Putting SMART Characteristics Together in Goals

For goals to work, they must have all five of the characteristics designated by *SMART*. The following example shows how this works: Administrators of a school district in a low-income area are very concerned about students’ relatively low attendance. A parental consultation program is created with a general goal of “increasing school attendance.” However, the goal is too general and provides little information on what is to be done, whom it will be done for or with. It also does not say how to measure progress toward the goal. With evaluators’ assistance, administrators develop the following two *SMART* goals:

SMART Goal 1

“By the end of 2014, at least 70% of parents with children having a low average attendance rate (65% of scheduled school days in the first semester of 2014) will have participated in the parental education program.”

The first goal is a *SMART* goal because it is . . .

Specific. It clearly indicates what will happen due to the program and who needs to participate (parents of children with low average attendance rate).

Measurable. It uses the parents’ participation rate (70%) as a criterion of success. The data can be collected by counting the target parents and those who attend.

Attainable. The goal of 70% was calculated by using parental participation rates in previous similar programs of around 60% to 70%.

Relevant. The goal is highly relevant to the purpose of the program, which must encourage parents’ participation in order to succeed.

Time-bound. The goal sets a clear deadline (the end of 2014) by which the criterion for success must be met.

SMART Goal 2

“By the end of 2014, children with parents participating in the intervention program will attend school 80% of scheduled days.”

The second goal is a SMART goal because it is . . .

Specific. It clearly indicates what will happen due to the program and who will engage in that behavior.

Measurable. It uses the rate of school attendance (80%) as an indicator of success. The data can be collected by counting all target children in the district and their days in school.

Attainable. The goal of an 80% attendance rate was set by using the baseline rate of 65%. Teachers have given feedback that the goal of 80% attendance is a bit of a challenge but attainable.

Relevant. The goal reflects the purpose of the program, an increase in school attendance.

Time-bound. The goal includes a clear deadline (the end of 2014) by which the criterion for success must be met.

EVALUABILITY ASSESSMENT

Evaluability assessment is a popular tool used to determine whether a rigorous outcome evaluation is warranted for a program (Smith, 1989; Wholey, 1987, 1994). If a program is regarded as having low evaluability, it does not meet the majority of the requirements raised by the above questions. If the program proves not to be evaluable, evaluators may be invited to work with stakeholders to foster its evaluability. Carrying out an outcome evaluation for a program that is not evaluable is not only imprudent but is also a waste of time and resources. Only if the evaluability assessment shows the program to be evaluable is an outcome evaluation then justifiable. Evaluability assessment has often been used as a program development tool for improving outcomes. In fact, it is very often a funding agency or program manager who recognizes that a program needs modification and initiates an evaluability assessment to make sure the program is an evaluable one.

In its basic form, an evaluability assessment assesses two general areas: program design and data collection capacity. More specifically, it attempts to answer the following questions:

- Does the program have an appropriate structure and design?
 - Does the program have a clear target population?
 - Does the program have a clear logic model or program theory?
 - Does the program have a set of specific and measurable goals?
- Does the program have the necessary resources to function well?
- Does the program have the capacity to collect data for an evaluation?
- Does the program collect baseline information on clients?
- Does the program collect data on implementation activities?
- Does the program monitor outcomes?
- Does the program have staff trained to collect data and track changes?
- Have the implementation and outcome data been used for program improvement?

Wholey (2004) proposed that an evaluability assessment follow six steps.

Step 1: Involve the Intended Users of Evaluation Information

Evaluators must plan to communicate with the program's stakeholders, from top to bottom, to ensure they are aware that the evaluability assessment is being conducted. Decision makers, managers, and program staff should understand that the evaluability assessment's purpose is to discover whether program designs conform to both the expectations of key stakeholders and the reality of program operation. Wholey (2004) stressed that the purpose of this step is to prevent evaluators from working in isolation, ignoring stakeholders' perceptions of the program.

Step 2: Clarify the Intended Program

Evaluators must study documentation of the program's history and funding and interview policy makers, managers, and staff to gather information about a program's intention. The purpose of this step is to clarify the relationship between a program's resources and activities and the intended outcomes that the resources and activities are expected to yield. Evaluators need to develop a logic model for the program and use it to describe the sequential relationships between inputs and activities and to explain how activities lead to the intended outcomes.

Step 3: Explore the Program's Reality

Evaluators must compare the logic model, which is primarily based on the vision of the stakeholders interviewed in step 2, and the program's actuality. Wholey (2004) indicated that to learn more about a program's reality, evaluators must examine documentation such as reports of accomplishments during the time period of interest and reports from past evaluations, as well as conduct site visits to obtain firsthand observation data. With this information, evaluators then compare the logic model to the program's reality. When discrepancies are found between the model and reality, evaluators must further identify the factors that inhibit program performance.

Step 4: Reach Agreement on Any Needed Program Changes

Evaluators sit down with key stakeholders to discuss what has been learned and what areas of the program need change. Evaluators use the information from step 3 about the extent of discrepancies between the logic model and program reality to discuss how likely the program is to succeed in achieving its intended outcomes. If discrepancies are significant, evaluators must recommend changes to the program's design and/or implementation before conducting an outcome evaluation. Areas of change may include resources, activities, and/or goals.

Step 5: Explore Alternative Evaluation Designs

If, based on the results of the evaluability assessment, decision makers want to move forward with a full-scale evaluation, evaluators can provide various design options for conducting the evaluation. For each potential design, evaluators may offer information about what data will be collected, how the data will be analyzed, the costs and time associated with the evaluation, and how the information gathered will be used.

Step 6: Agree on the Evaluation's Priority and How Information From the Evaluation Will Be Used

Evaluators and managers must outline an agreement for evaluating the program. The agreement needs to prioritize which program areas are to be evaluated in what time frame and how the information will be used.

PLAUSIBILITY ASSESSMENT/CONSENSUS-BUILDING APPROACH

The plausibility assessment/consensus-building approach is used to identify plausible intended and unintended outcomes as a basis for facilitating stakeholders in building consensus around a set of plausible goals for management and evaluation purposes. This section is an expansion of Chen's (1990) work. Traditionally, evaluators tend to use official goals as a basis for developing a set of measurable outcomes to use in assessing a program's effectiveness. However, this approach has two potential problems: (1) It may not identify what a program is really doing and overlook important effects, and (2) stakeholder groups may not agree on which goals the program should be evaluating. These problems prevent a program from functioning effectively. This section will first discuss the issues that arise when an evaluator uses official goals uncritically to conduct an evaluation. It then introduces the plausibility assessment/consensus-building approach, which can address these problems.

Potential Problems of Evaluation That Are Based Mainly on Official Goals

Every intervention program or policy has a set of official goals stating what is to be accomplished. These official goals are formally stated in a program's formal documents such as memorandums, grant applications, brochures, fliers, and so on. One of the main reasons evaluators may use these goals as the basis for assessing a program's effectiveness is that using them appears to provide legitimacy to the evaluation, as the official goals were often developed and are supported by the program's leadership. However, two problems may arise when evaluators use official goals as the criteria for evaluating a program's effectiveness. The first is the "goal trap" problem. The second is that stakeholder groups may not agree on which goals or outcomes should be evaluated.

The Goal Trap

A goal trap occurs when evaluators (1) fail to recognize that some official goals are window dressing and (2) neglect to look into program activities and effects not covered by the official goals.

1. Official goals may be established to serve political rather than evaluation purposes. These goals are usually stated to project an appealing image to the public rather than to reflect the reality of what the program does (Chen, 1990). For example, a youth program may have a goal of enhancing participants' self-esteem—surely an unobjectionable aim—but not engage in specific activities or services that could reasonably link the program to such a goal. An evaluation

of window-dressing goals is likely to show program failure, and would be a waste of money and other resources.

Sometimes, program management and staff may list official goals that they want to pursue in the future but currently do not have the resources to work toward. Again, evaluators should be sensitive to this issue and communicate with stakeholders about whether to include them in an evaluation. In this way, evaluators can avoid the goal trap problem.

2. By focusing entirely on official goals, evaluators pay little attention to what the program really does. Official goals may blind evaluators to investigating actual effects. Scriven (1972) argued that official program goals could lead evaluators to narrowly confine evaluation activities to goal-related activities and pay little attention to actual effects. To address the problem, Scriven proposed *goal-free evaluation*. Goal-free evaluation is carried out when evaluators are not exposed to, or contaminated by, knowledge of program goals. Scriven argued that “the less the external evaluator hears about the goals of the project, the less tunnel-vision will develop, the more attention will be paid to *looking for actual* effects (rather than *checking on alleged* effects)” (1972/1991b, p. 57). Whether goal-free evaluations are feasible is not clear, but Scriven made an important contribution in calling evaluators’ attention to the need to take extra care when using official goals as a basis for evaluation.

Goal trap problems tend to be more problematic for programs, policies, or institutions with broad aims or large scope, such as Head Start, youth services, community health centers, prisons, detention centers, welfare programs, and so on.

Disagreements About Which Goals to Evaluate

Another problem with taking official goals at face value is that stakeholders may agree on which goals should be stated in the program’s documentation but not agree on which goals should be included in an evaluation. It is one thing to reach consensus on a set of program goals to list in a brochure so as to enhance a program’s public appeal. It is another thing to agree on which goals should be selected for evaluation. Some stakeholders may favor using official goals to do evaluation; others may not. For example, decision makers, who have much more authority and power than the supervisors and staff who actually deliver services to clients, tend to believe that using the official goals as a basis for evaluation is a reasonable way to proceed. Other stakeholders may have reservations about such an approach.

These problems are especially severe in large-scale programs. For example, in the current Veteran Affairs (VA) scandal, the VA administration set as one of the goals for VA hospitals as achieving a maximum 14-day waiting time for new appointments. Many managers in VA hospitals viewed this goal as unrealistic. Unfortunately, there was no mechanism available for the top administration and

hospital managers to candidly discuss the issue. As a result, it led to false reports and hidden wait times in many hospitals because managers' performance and salary increases were tied to the goal attainment. The VA's internal audit found 57,000 veterans waiting for up to three months for medical appointments and 64,000 who enrolled for VA health care over the past decade who had never been seen by a doctor.

Another example to illustrate the problem of disconnection between stated program goals and actual actions is the first evaluation of Head Start (Smith & Bissell, 1970). Evaluators based their assessment of its effectiveness mainly on its official goals, which included increasing disadvantaged preschool children's reading and math scores. The evaluation results indicated that the program was not effective. Local Head Start directors and teachers protested the evaluation by pointing out that it did not assess what they were actually doing. They viewed Head Start as a program that took a holistic approach to helping disadvantaged children—improving their nutrition and dental health, communication skills, emotional growth, and community engagement. Head Start staff also taught the parents parenting skills. Math and reading scores were a minor outcome among these various objectives. Evaluators, on the other hand, defended their evaluation as based on the program goals stated in the program's official documentation and endorsed by decision makers. In this case, obviously, decision makers, local directors, and teachers disagreed about which goals should be evaluated. The evaluators' insensitivity to this issue resulted in controversy.

Plausibility Assessment/Consensus-Building Approach

The plausibility assessment/consensus-building approach can address the two issues raised above: the goal trap and disagreements about which goals should be the basis for evaluation. This approach is useful for drawing evaluators' attention to the nature of program goals and their relationship to evaluation, improving the likelihood that an evaluation will detect actual effects, increasing stakeholders' buy-in to the outcome evaluation, and ensuring the usefulness of the evaluation.

This approach consists of five phases. The following sections will introduce the theoretical background supporting each phase and the procedures an evaluator can follow in conducting each phase.

1. Identify Which Official Goals Are Plausible and Which Are Not

As discussed earlier, not all official goals are suitable for evaluation purposes. It wastes both money and effort to evaluate goals intended as window

dressings, as such an evaluation is unlikely to find that the program is having its alleged intended impact. Such a finding often stirs up controversy. Plausible goals are official goals that a program is actually pursuing and on which the intervention is likely to have an impact.

Methods for Differentiating Plausible Goals From Window-Dressing Goals

- a. Sufficient resources are allocated to implementing activities related to the goals.

One criterion to differentiate plausible from implausible goals is resource allocation. As discussed previously, agencies may claim to have certain goals in order to please various coalitions or for other political purposes. An examination of resource allocation and operative activities can make clear which goals are plausible and which are not. For example, it would be difficult for administrators to claim rehabilitation of criminal offenders as a plausible goal if few resources or activities are directed at rehabilitation.

For example, a fair-housing contract service in a community had a mission to prevent housing discrimination. The agency formally stated nine goals in areas such as affirmative action and marketing, assessment of the housing-delivery system, complaint investigation, metro area outreach, education, interagency coalitions, referrals and exchanges, and making contact with home seekers. Records of expenditures showed that of these nine goals, only five received the vast majority of the available resources: affirmative action and marketing, assessment of the housing-delivery system, complaint investigation, metro area outreach, and education. The agency had not seriously pursued the other four goals, at least not at the time of the evaluation. Subsequent interviews with the program manager and staff, and observation of the program's operation, confirmed the initial impression.

- b. The goals are consistent with the existing knowledge, experience, and/or understanding of the problem to which the program is directed.

A plausible goal will have at least some connection with prior knowledge or experience. For example, a local television station launched three hour-long episodes of a program on the roots and consequences of racism. According to the station, the official goal of the broadcast was to eliminate racism in the community. However, the existing literature on racism indicates that eliminating racism is extremely difficult. It is therefore highly implausible that the broadcast would eliminate racism in the community. An evaluation of this official goal would show the program's seeming ineffectiveness and waste money. Evaluators could suggest that stakeholders

consider using a more plausible goal, such as increasing residents' awareness of the adverse consequences of racism, as the basis of evaluation.

c. Implementers report performing activities consistent with the goals.

Implementers have first-hand information about which of the official goals they are or are not pursuing. Evaluators can ask them about their weekly or monthly activities related to these goals. In these interviews, evaluators must provide implementers with a safe environment in which to express their views.

2. Identify Operative Goals That Do Not Appear in the Official Goals

The second step of the plausibility assessment/consensus-building approach is to identify whether any operative goals are not listed as official goals but are in fact pursued by implementers and staff (Chun & Rainey, 2005; Perrow, 1961). According to Perrow, official goals are "the general purposes of the organization as put forth in the charter, annual reports, public statements by key executives, and other authoritative pronouncements" (p. 855). He defined operative goals as "the ends sought through the actual operating policies of the organizations, they tell us what the organization is actually trying to do, regardless what the official goals say the aim is" (p. 855). Because official goals may reflect only a desirable state of affairs, rather than the realistic outcomes to which an organization's members are committed, Perrow argued that organizational effectiveness can be better understood by studying operative goals than official goals.

The distinction between official goals and operative goals is especially applicable to social betterment and health promotion programs, which have been found to have both kinds of goals. For example, correctional institutions usually claim rehabilitation of prisoners as an official goal, but in fact these institutions mainly pursue custodial care and deterrence as operative goals. The official emphasis on rehabilitation is to make the program more appealing and acceptable to the public. Similarly, a community-based program may list an official program goal of enhancing the well-being of people with disabilities. However, its real operative goal may be to raise funds for community activities by selling products made by people with disabilities.

When planning an intervention program or policy that decision makers, or funding agencies, have high hopes for, program decision makers often do not know exactly how to translate these hopes or goals into actions. The work of defining target groups, recruiting or screening applicants, communicating with clients, allocating resources, and providing services and so forth are left to program managers and implementers (Chen, 1990). Since program managers

and implementers exercise a large degree of discretion in the day-to-day program activities that shape the program's actual direction, the intervention delivered in the field may not align with what was originally described in the program goals.

Since official program goals do not necessarily include operative goals, an evaluation narrowly focused on official program goals is bound to neglect the real effects of the program and provide a partial or even biased evaluation of the intervention's effectiveness. The methods used to collect data for identifying plausible goals and window-dressing goals, discussed above, can also be used to identify operative goals.

Methods for Identifying Operative Goals

a. Inquire into resource allocation.

An examination of how the program allocates its resources provides clues to intended outcomes. For example, in Zald's (1963) study of correctional institutions for delinquents, the official goal was rehabilitation. Zald examined resource allocation within the programs to see whether this official goal was actually pursued. Information pertaining to resource allocation was obtained from both the observation of daily work and interviews with program managers and staff at the correctional institutions. Zald discovered that resources were mainly allocated to the custodial aspects of the organizations (operative goal) rather than to professional rehabilitative efforts (official goals). The data clearly indicated that custody was an operative goal of these organizations that they pursued to a greater extent than the official goal of rehabilitation.

b. Inquire into eligibility, recruitment, and service delivery processes.

An investigation into eligibility for and recruitment to the intervention and the program's service delivery processes may indicate the operative goals. In a study of the Ohio District Eleven Adoption Project, Olsen (1981) observed that the original goal of the project was to serve children with disabilities. However, after implementation, it was found that a much broader group needed services. Children who were racial minorities were overrepresented in the backlog of cases. Information from the profiles of children who needed services led to a change in the program goal from serving only children with disabilities to the broader goal of meeting the needs of minority children. Olsen (1981) noted that, in spite of the project's expansion of purpose during the period of operation, its official goals were never changed.

c. Interview implementers.

Again, evaluators can examine the weekly or monthly activities of the staff members delivering the intervention and see whether these activities are consistent with the official goals. If not, this is an indicator that operative goals should be considered.

3. Identify Possible Unintended Outcomes

An intervention program is not a goal machine. As discussed in Chapter 1, it operates in a social system and interacts with other components in the system and its environment. As such, it likely generates unintended outcomes, which may be positive or negative, and outcome evaluation needs to address these outcomes. An example of a negative unintended outcome is a food aid program whose official purpose is to reduce acute and chronic food insecurity of individuals and communities. However, the program may have undermined recipients' capacity to meet their own basic needs without external assistance. Another example of a possible negative unintended outcome is sex education in schools, which, critics argue, may encourage youth to experiment with sex.

However, it is important to realize not all unintended outcomes are negative. For example, a daylong summer tutorial program for children of Vietnamese and Laotian descent may not only improve their school grades as intended but also, by having children of these two ethnic groups learn and play together, may reduce ethnic feuds between these two groups. In another example of a positive unintended outcome, the initial purpose of the 55-mile-per-hour speed limit was to save fuel in response to the Arab oil embargo and subsequent high oil prices. However, researchers found that the policy also reduced fatalities, injuries, and property damage (Clotfelter & Hahn, 1978).

If a program has important unintended outcomes, an evaluation that does not first identify these may not provide a fair assessment. Little of the literature discusses how to address unintended outcomes. Perhaps it is challenging to identify unintended outcomes before program implementation for evaluation purposes. Evaluators can use the following two methods to identify and study unintended outcomes.

Methods for Identifying Unintended Outcomes

- a. Examine a program's eligibility requirements, recruitment process, and service delivery.

An examination of a program's eligibility requirements and recruitment process, as well as its service delivery, may provide clues to unintended outcomes. For example, the official goal of Aid to Families with Dependent Children (AFDC) was to relieve economic hardship in needy families. To target the program at the most needy populations, eligibility requirements restricted benefits to female-headed families. This eligibility requirement could lead many parents to choose not to marry or live together or many husbands to stay at home rather than work so their families could qualify for benefits. The same thing could happen in service delivery. For example, an evaluation workshop for community-based organizations might provide an opportunity for participants to network and develop the basis for collaboration on future projects.

If stakeholders are concerned about, or interested in knowing, about plausible unintended outcomes, evaluators need to work with them to identify such outcomes.

b. Use existing theory or knowledge.

Existing theory and knowledge provide another means for identifying potential unintended outcomes. For example, Waldo and Chiricos (1977) applied several theories from criminology to identify various unintended outcomes of a work-release program. Based upon these theories, they argued that the effects of a work-release program may go beyond reducing recidivism (official goal) by helping participants to develop responsibility and self-discipline and increasing opportunities for social adjustment (unintended outcomes). Waldo and Chiricos argued that assessing program effectiveness by measuring recidivism alone could overlook other important benefits.

4. Discuss the Findings of the Plausibility Inquiry With Stakeholder Groups to Obtain Input, Commentary, and Decisions

Evaluators need to draft the findings of the plausibility inquiry, discuss them with stakeholders, seek their input, and revise the findings based on their comments. As discussed previously, different stakeholder groups, such as decision makers and implementers, may have different views on which goals and outcomes should be included in the evaluation. Evaluators need to discuss the draft with representatives of each group initially. There are important reasons to meet with different groups separately. Different stakeholders have different levels of power and authority. If evaluators invite them to the same meeting,

those with low power and authority usually are quiet or do not express their true feelings. For example, if decision makers and implementers are both in a meeting, implementers are less likely to state their opinions openly.

An effective strategy for dealing with unequal power situations is to hold separate meetings with representatives of each group. The evaluator can then facilitate stakeholders in reaching consensus on a set of goals and unintended outcomes for evaluation purposes.

5. Revise and Finalize the Set of Program Goals and Outcomes for Evaluation

To ensure that stakeholder groups agree on the set of goals/outcomes, it is crucial for evaluators to send the revised draft to stakeholder groups for another round of comments and suggestions before finalizing it. The purpose of this feedback process is to help stakeholders come to agreement on a set of plausible goals and outcomes for evaluation purposes. The revised set of program goals and outcomes is usually presented in a large meeting with representatives of different stakeholder groups for final edits and approval. An iterative process of seeking input from stakeholders to the development of a set of goals and outcomes for evaluation purposes increases stakeholders' support of the evaluation and their utilization of evaluation results.

A PREVIEW OF CONCLUSIVE OUTCOME EVALUATION: SELECTING AN APPROPRIATE APPROACH

When a program is ready for a conclusive outcome evaluation, evaluators and stakeholders can discuss and plan the evaluation. This book will introduce two approaches that relate to conclusive outcome evaluation. Chapter 10 will discuss the experimentation evaluation approach to conducting conclusive outcome evaluation. This traditional approach seeks to maximize internal validity (rigor) in the assessment of an intervention's effects on outcomes. Chapter 11 will introduce the holistic effectuality evaluation approach to assessing an intervention's real-world effects. This approach proposes a hybrid outcome evaluation that includes both constructive and conclusive assessments. Evaluators must be familiar with and competent in both areas and discuss each approach's strengths and limitations with stakeholders so that an appropriate approach can be chosen to guide the design of a conclusive outcome evaluation.

QUESTIONS FOR REFLECTION

1. What is the purpose of a constructive outcome evaluation?
2. If a program has clear goals, is it still necessary to employ a constructive outcome evaluation? Explain.
3. Define the components of SMART goals. Develop the SMART goals method for the following goal statement of an elderly service program: “enhancing senior citizens’ quality of social life.”
4. How does the evaluability assessment determine whether a program requires a rigorous outcome evaluation? What enables an evaluability assessment to facilitate program development?
5. Discuss the steps provided by Wholey (2004) for an evaluability assessment. How do they relate to the SMART goal components?
6. Discuss the repercussions of a goal trap.
7. Discuss the merits of the goal-free evaluation and the feasibility of conducting a goal-free evaluation in a real-world setting.
8. Discuss the reasons why stakeholder groups may not agree on which goals should be evaluated. Give some examples.
9. What methods can be employed to determine whether or not goals are plausible?
10. How do operative goals differ from official goals? Give examples.
11. Discuss how to identify plausible intended outcomes and plausible unintended outcomes.
12. Discuss how evaluators could facilitate stakeholder groups in reaching agreement on which goals will be evaluated.
13. Discuss why managers in local VA hospitals did not immediately bring to their superiors’ attention the fact that the goal of a maximum 14-day waiting time was unattainable. In your opinion, if evaluators had been invited to evaluate the program before the scandal broke, could they have facilitated the different levels of management in discussing and addressing the problem of unrealistic goals? Why?

CHAPTER 10



The Experimentation Evaluation Approach to Outcome Evaluation

This chapter introduces the current dominant approach to conclusive outcome evaluation: the experimentation evaluation approach. Discussed here are the concepts, theory, and methodology of this approach and how to use experimentation evaluation to design conclusive outcome evaluations. As will be discussed later, the focus of the experimentation evaluation approach is on ensuring the internal validity (rigor) of outcome evaluation. Accordingly, outcome evaluation guided by this approach is called in this book *validity-focused outcome evaluation*. This approach has been applied in research or real-world settings. An alternative approach to outcome evaluation, the holistic effectuality evaluation approach, will be discussed in the next chapter.

THE FOUNDATION OF THE EXPERIMENTATION APPROACH TO OUTCOME EVALUATION

When program evaluation was formally introduced as an applied science in the 1960s, evaluators looked to conceptual frameworks and methodologies for guidance on how to evaluate the effectiveness of an intervention program. Suchman (1967) was convinced that the concepts, principles, and methods proposed by the Campbellian validity typology (Campbell & Stanley,

1963) for experimental research could serve outcome evaluation purposes very well. He argued that by following the Campbellian validity typology, evaluators could provide rigorous and credible evidence as to whether an intervention was effective. Later, many evaluators joined the movement, also advocating use of the Campbellian validity typology (Campbell & Stanley; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002) as a foundation for designing and conducting outcome evaluation. This movement is so popular that major evaluation textbooks (e.g., Posavac, 2011; Rossi et al., 2004; Wholey, Hatry, & Newcomer, 2010) devote chapters to discussing how to apply the typology in conducting outcome evaluation.

The experimentation evaluation approach is defined in this book as an advocacy for the application of the principles, theory, and methodology of the Campbellian validity typology to designing and conducting outcome evaluations. Most evaluators, including the author, were taught in their evaluation courses to apply Campbellian validity typology, and this tradition continues to be popular.

THE DISTINCTION BETWEEN INTERNAL VALIDITY AND EXTERNAL VALIDITY IN THE CAMPBELLIAN VALIDITY TYPOLOGY

Evaluators need concepts and principles to help them better understand what kinds of evidence need be included in an outcome evaluation and how to gather them. The experimentation evaluation approach uses the distinction of internal and external validity (Campbell & Stanley, 1963) to categorize evidence. According to Campbell and Stanley, internal validity asks the question “Did the experimental treatments make a difference in this specific experimental instance?” In the case of evaluation, *internal validity* means the extent to which a study can provide accurate information about whether an intervention has produced observed effects on outcomes. According to Campbell and Stanley, external validity asks a question about generalizability such as “To what populations, settings, treatment variables, and measurements can this effect be generalized?” In the case of evaluation, *external validity* means whether or not the evaluation results can be generalized to these areas.

Cook and Campbell (1979) and Shadish and colleagues (2002) have revised the internal and external classification of validity. One important feature of their revision is a subdivision of internal and external validity into two additional categories: *statistical conclusion validity* and *construct validity*. The former refers to the appropriateness of drawing conclusions from the statistical evidence. The latter refers to making generalizations about higher-order constructs from the research. However, the evaluation community still makes wide use of Campbell and Stanley’s original categories of internal and external validity.

Campbell and Stanley (1963) indicated that both internal and external validity are important for a study, but there is trade-off between them: An increase in internal validity in a study leads to a decrease in external validity, and vice versa. In addressing this trade-off, Campbell and Stanley forcefully argued that the prime priority of a study is internal validity. According to them, “internal validity is the basic minimum without which any experiment is uninterpretable” (p. 3). This view provides Campbell and his associates a justification for focusing attention more heavily on internal validity than on external validity. The experimentation evaluation approach has accepted this view, and evaluators have focused on addressing internal validity issues when conducting a conclusive outcome evaluation.

THREATS TO INTERNAL VALIDITY

According to the Campbellian validity typology, a research design is rigorous when it is capable of ruling out potential rival hypotheses or confounding factors related to an intervention or treatment. This typology stresses that it is essential for research to provide evidence that the change in an outcome is attributable to an intervention or treatment rather than to rival hypotheses or confounding factors. The argument is persuasive, but an issue arises when one tries to determine what the rival hypotheses look like and how to address them in a study. The Campbellian validity typology makes a profound contribution to the social and behavioral sciences by classifying these rival hypotheses and confounding factors, as well as illustrating what they look like. This perspective will be the focus of the following section. Campbell and associates foresaw these rival hypotheses, or confounding factors, as “threats to internal validity” and defined them as follows:

History. An event other than the intervention may happen before or during the implementation of that intervention that could affect an outcome. For example, an evaluation of the effectiveness of a HIV-prevention education program, targeted at youth, may be conducted at the same time that a popular movie dramatizes the consequences of unsafe sex. In this case, the movie could be a threat to the internal validity of the evaluation.

Maturation. Participants may naturally change due to the passage of time (e.g., growth or fatigue), and such change is generally not a result of an intervention. For example, a cold medicine may cure a person’s cold, but the person’s immune system could have overcome the cold in a week had the person not taken any medicine. In this case, the person’s immune system is a threat to evaluating the effectiveness of the medicine.

Testing. Participants who have completed a pretest may learn from that experience and thus do better on a posttest, regardless of the effects of an intervention. For example, people participating in a job-training program may perform better on a job interview as a result of learning how to present themselves from previous interviews, rather than from the job-training program.

Instrumentation. The measurement instrument itself may change over time. For example, drug arrest rates in a community may decrease because police officers are using different criteria to enforce the law, rather than due to a community-wide drug abuse prevention program.

Statistical regression. Extreme scores, those that are much higher or much lower than the mean, tend to move toward the mean score upon subsequent retesting. For example, if students who scored in the 10th percentile on a verbal test take another verbal test, their scores on the second test will tend to be better than on the first test, even without any intervention. Since these students scored extremely low to begin with, we would expect their second test to be better than the first.

Attrition. A program may lose participants because they drop out, get sick, etc. For example, high-performing participants who leave the intervention group because they feel they no longer need the treatment could affect posttest scores.

Selection. Nonrandom factors may influence participants' selection of or inclusion in different groups. For example, highly motivated people may join the intervention group, while less motivated people may join the comparison group. These group differences could affect posttest results even without an intervention.

Selection interactions. A selection threat may combine with another threat of internal validity to confound analysis of the intervention variable. For example, two groups may be different in terms of initial performance (selection), and the high-performance group also matures faster than the low-performance group (maturation).

RESEARCH DESIGNS FOR RULING OUT THREATS TO INTERNAL VALIDITY

Following the Campbellian validity typology's principles, the experimentation evaluation approach argues that evaluators must use research designs to rule out threats to internal validity in an evaluation. In this way, they can provide

credible evidence of a causal relationship between an intervention and outcomes. This approach argues that by ruling out threats to internal validity, an outcome evaluation can provide credible evidence that it is the intervention, not threats to internal validity (confounded factors), that has affected outcomes. Thus, this approach heavily emphasizes the assessment of an intervention's pure independent effects on outcomes. However, not all research designs are equal in accomplishing this task. According to the Campbellian validity typology, pre-experimental designs, such as surveys, are the weakest designs because they cannot rule out most of the threats to internal validity. The strongest designs include randomized experimental designs, which can rule out most or all threats to internal validity. Quasi-experimental designs, while better than pre-experimental designs, are not as strong as experimental designs.

The rest of this chapter will discuss three types of designs: experimental, quasi-experimental, and pre-experimental. The Campbellian validity typology provides a set of symbols to describe various designs: An *X* represents an intervention. An *O* refers to an observation of outcome before or after an intervention. An *R* stands for the random assignment of participants to different research conditions. These symbols are easy to understand and will be used in this chapter to discuss the structure of research designs.

Experimental Designs

According to the Campbellian validity typology, the strongest designs for providing credible evidence are experimental designs. Typically, experimental designs are diagrammed as follows:

Intervention Group:	R:	O_1	X	O_2
Control Group:	R:	O_1		O_2

In this kind of design, participants are randomly assigned to a treatment group or control group. The core of experimental designs is random assignment, which creates two equivalent groups for comparison. The Campbellian validity typology argues that random assignment eliminates many threats to internal validity. For example, participants in a research project may be different in terms of history of sickness, gender, weight, and so on, and these variables may correlate with the intervention. Researchers deal with this problem by randomly assigning clients to the experimental or control groups. Randomization generates equivalent experimental and control groups for an unbiased assessment of intervention effect.

Among experimental designs, the crown jewel is randomized controlled trials (RCTs), also known as efficacy evaluations (Flay, 1986). In addition to using random assignment to place participants in an experimental group or control group, RCTs further strengthen the rigor of the study in the following ways: (1) using homogenous groups to reduce confounding factors, (2) following strict guidelines when recruiting study participants, (3) demanding high fidelity when implementing intervention protocols, and (4) utilizing a double-blind technique (i.e., neither participants nor counselors/implementers know which participants are in an intervention or control group) or a triple-blind technique (i.e., even the researchers do not know who was in an intervention or control group).

RCTs have been applied widely and successfully in biomedical research and have made profound contributions to medicine. Both researchers and academics view RCTs as the gold standard of scientific research, a view shared by proponents of the experimental evaluation approach. In addition to being scientifically rigorous, RCTs are also popular for the following reasons:

- Panels that review grant proposals for the National Institutes of Health (NIH) and many other federal funding agencies, favor RCTs or efficacy evaluation, because these approaches exhibit scientific objectivity.
- Reputable journals prefer to publish articles characterized as well-controlled efficacy studies.
- By creating ideal research conditions, researchers can control participants and other elements of the study, making evaluation and research much more manageable.
- As discussed at the beginning of this chapter, the belief is common (and strongly held) among scientists that an intervention's efficacy needs to be evident *before* any examination of its effectiveness in the real world or generalizability can have meaning.

Case Illustration of the Experimental Design

The following case study is a well-known evaluation of a school-based health education program for children with asthma, aged 8–11 years, that used efficacy evaluation (Evans, Clark, & Feldman, 1987). The program, Open Airways, consisted of six, 60-minute sessions in which groups of 8 to 12 students learned new asthma management skills. The program was taught by health educators. The sessions included topics such as how to recognize and respond to symptoms of asthma, how to use asthma medicines and decide when to seek help, how to stay physically active, and how to identify and control triggers of asthma symptoms.

The program focused on training children to be self-managers, emphasizing the children's responsibility for recognizing asthma symptoms and initiating management steps whether or not a parent was present. The intervention was designed to enhance children's feeling of self-efficacy, with health educators verbally encouraging children to believe in their own ability to manage wheezing episodes and—through modeling, practice, and rehearsal—helping children to recognize that they could successfully perform the new asthma management skills. The six program sessions were held during school hours and were offered over a 2- to 3-week period. If necessary, makeup sessions were held to ensure that children completed the entire program.

Twelve public elementary schools from two school districts in the community participated in the program. Parents who knew or believed that their children had asthma and wanted them to participate in the program were invited to participate as well. Parents were telephoned or interviewed to determine whether the child met the criteria for participation. A total of 239 children took part in the study. After enrollment, the 12 schools were paired according to ethnic composition and size. In each pair, one school was randomly assigned to receive the health education program, while the other school served as a control group and did not receive the program.

Data about asthma management skills, feelings of self-efficacy, parents' management decisions, school absences, and school performance were collected from the child's school records, from the medical records of hospitals where emergency care was sought, and from separate interviews with parents and children before the intervention and 1 year after its completion. Data were analyzed by using analysis of covariance. The study found that, when compared to the control group, children in the experimental group had higher scores on asthma management, greater self-efficacy with regard to asthma management skills, more influence on parents' asthma management decisions, better grades in school, and fewer episodes of asthma. The program had no effect on school absences.

Pre-Experimental Designs

Pre-experimental designs do not resemble experimental designs, and the Campbellian validity typology argues that these are the weakest designs for assessing an intervention's effects. This section will introduce three pre-experimental designs.

One-Group Posttest-Only Design

This design only measures the outcome after the intervention.

Intervention group: X O

The Campbellian validity typology regards this design as the weakest because it lacks baseline data and cannot rule out most of the threats to internal validity. For example, a parenting program offered to mothers of disadvantaged children might anticipate that it would have a positive influence on parental involvement in children's schoolwork and other activities. After the intervention, the mothers' parenting knowledge and skills would be tested. However, without baseline information pertaining to their prior knowledge and skills, it would be difficult to determine whether the intervention had impacted the intended outcomes.

Static Group Comparison

This design expands on the one-group posttest-only design by adding a comparison group. The design is diagrammed as follows:

Intervention group: X O

Comparison group: O

The dotted line between these two groups signifies that they are not equivalent. By adding a comparison group, the static group comparison design is slightly better than the one-group posttest-only design. Survey research is based upon this design. One of the major limitations of the design is a lack of baseline information. For example, the mothers attending the parenting class may come from one neighborhood, while parents in the comparison group may come from another neighborhood. If the posttest shows that mothers in the intervention group have better parenting skills than mothers in the comparison group, it is difficult to determine whether the difference is due to the intervention or due to preexisting differences.

One-Group Pretest-Posttest Design

Another pre-experimental design is the one-group pretest-posttest design. The design is diagrammed as follows:

Intervention Group: O₁ X O₂

This design upgrades the one-group posttest-only design by adding a pretest component. However, according to Campbellian validity typology, this design is still very weak because it does not rule out most of the internal validity threats. For example, history is one potentially uncontrolled variable; other parenting skills interventions may have taken place in the community. Testing is another potential problem, as the pretest could sensitize parents to the skills and knowledge central to the intervention and thus result in better performance on the posttest. Maturation is yet another potential problem; the differences between pretest and posttest may result from a community or national trend toward greater parental involvement in children's schoolwork and other activities.

It is interesting that the one-group pretest-posttest design, despite its weaknesses, is frequently used by community-based organizations and even by researchers. We will revisit this design in the next chapter.

Case Illustration of the Pre-Experimental Design

The Reinforcement Alcohol Prevention (RAP) Program (Will & Sabo, 2010) provides a case study illustrating the application of the one-group pretest-posttest design. RAP applied social cognitive theory (SCT) to translate personal, behavioral, and environmental factors into a school-based program. The program addressed teens' knowledge and perceptions of alcohol and raised awareness of the harmful side effects of alcohol. The program enrolled 8th-, 9th-, and 10th-grade students from 55 middle school and high school classrooms. RAP was delivered to students in small- and large-group settings during health and physical education classes. The intervention included packaged curricula, presentations, incentivized learning activities, and threat-appeal tactics such as photographs of automobile accidents and personal testimonials. The program focused on equipping teens with practical ways to say no to alcohol and bolstering self-efficacy. It provided accurate information about rates of drinking among participants' teenaged peers and scenarios that enabled teens to practice refusing an offer of alcohol. The curriculum consisted of 1-session, 90-minute modules. Each module included a student education component, a 5-minute video about a car crash victim, and interactive activities.

Outcome measures included a 1-page, 25-item questionnaire, and 668 students were asked to complete the survey before and after the intervention. The data indicated that students' knowledge and awareness of alcohol-related risks and consequences increased from the pretest to the posttest. The authors concluded that this preliminary study showed the program was well received and significantly improved students' knowledge and awareness of alcohol's harmful effects.

Quasi-Experimental Designs

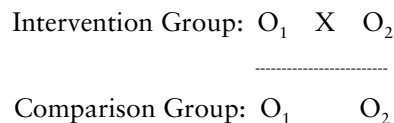
According to the Campbellian validity typology, quasi-experimental designs have features similar to experimental designs but do not use random assignment. Quasi-experimental designs do not remove all of the threats to internal validity, but they remove the majority of them. Campbell and his associates stated that these designs are not as strong as experimental designs but are much superior to pre-experimental designs.

Quasi-experimental designs are a popular choice for evaluations. Evaluators often find it is very difficult to apply RCTs or other experimental methods because they do not meet ethical or feasibility criteria. Evaluators then use quasi-experimental designs, despite anxiety about the designs' potential limitations. When applying the quasi-experimental design, evaluators use different techniques in an attempt to limit threats to internal validity and enhance the rigor of the design. For example, one technique is to select a comparison group that closely resembles the intervention group. Furthermore, it is also common for evaluators to suggest that future research use RCTs or other experimental methods.

This next section will discuss two popular quasi-experiments: nonequivalent comparison group design and interrupted time-series design.

Nonequivalent Comparison Group Design

The design is diagrammed as follows:



This design has two groups: an intervention group and a comparison group. The dotted line between these two groups indicates that the intervention and comparison groups are not exactly equivalent. In this design, the intervention group is observed, subjected to the intervention, and then observed again. The second group is a comparison group that undergoes a pretest and posttest but not an intervention. In other words, by adding a comparison group to the one-group pretest-posttest design, researchers upgrade to a quasi-experimental design. Using the parenting-involvement program as an example, this could mean adding a group of parents of disadvantaged children in a nearby community with similar demographics to be used as a comparison group. The group has a pretest

and posttest given, just as does the intervention group, but the parents in this group do not receive education or counseling on how to help their children with schoolwork and other activities. The difference between this design and an experimental design is that parents are not randomly assigned to the two groups.

According to Campbellian validity typology, this design can eliminate threats such as history, testing, and maturation. The results are often more interpretable than those of pre-experimental designs.

Case Illustration of a Quasi-Experimental Design

Nonequivalent Comparison Group Design. Dutton (1986) applied a nonequivalent comparison group design to assess the effectiveness of a court-mandated treatment for spouse abuse. The treatment group included 50 men convicted of spousal abuse. The comparison group also included 50 men convicted of spousal abuse, but these men were not treated by the program. The treatment consisted of 4 months of court-mandated group therapy, which included cognitive behavioral modification, anger management, and assertiveness training. Participants in the treatment group met in groups of eight for 3 hours each week. Furthermore, outside group exercises and buddy system hotlines were established. At the end of the 4 months, three optional couple communication therapy sessions were available for the men to attend with their spouses. Both groups were surveyed using Straus's Conflict Tactics Scale (CTS) before and after treatment. The clients' wives were also asked to independently fill out a CTS before and after the intervention. In addition, data about repeated assaults were gathered for both groups from police records. The treated men and their spouses had lower CTS scores than did the untreated men and their wives. Furthermore, the treatment group had a 4% recidivism rate during a posttreatment period of up to 3 years, while the comparison group had 40% recidivism during the same period. The researchers believed the program to be a success, but they also expressed that caution be used when interpreting the data due to the limitations of quasi-experiments, and they recommended that a randomized experiment be used for future studies.

Interrupted Time-Series Design. The interrupted time-series design is characterized by multiple observations of an outcome variable for a unit, such as an organization or community, before and after an intervention. The design is diagrammed as follows:

$$O_1 O_2 O_3 O_4 O_5 X O_6 O_7 O_8 O_9 O_{10}$$

Again, this design can be seen as an expansion of the one-group pretest-posttest design, because it adds more observations before and after an intervention. An example of the application of this design is a city government that wanted to reduce juvenile violent crime by initiating a citywide police-probation partnership (Worrall & Gaines, 2008). Police officers were paired with probation officers to enhance the supervision of juveniles who were arrested. The supervision included unscheduled home and school visits to ensure that juvenile probationers were in compliance with their probation requirements such as obeying curfews, paying fines, and attending school. By applying an interrupted time-series design, evaluators collected the monthly number of juvenile violent crimes for a period before and after the intervention. Results of this study showed a citywide reduction in assault, burglary, and theft arrests. According to the Campbellian validity typology, the interrupted time-series design can address all of the threats to internal validity except history.

QUESTIONS FOR REFLECTION

1. Define the experimentation evaluation approach. Discuss why this approach is so popular in evaluation.
2. Discuss internal and external validity and their meaning for evaluation. How can each be assured in the experimentation evaluation approach?
3. Following the Campbellian validity typology, the experimentation evaluation approach argues that internal validity should be the prime priority in evaluation. Do you agree or disagree? Why?
4. Give two examples of each threat to internal validity.
5. Why is it important to rule out the threats to internal validity? What could be the consequences if an evaluator is unable to rule out all or most of the threats to internal validity?
6. Why are experimental designs considered the strongest research designs? Do they have limitations in the context of evaluation and, if so, what are these limitations?
7. Why are randomized control trials (RCTs) considered the crown jewel of experimental designs? How do they rule out threats to internal validity?
8. Find an evaluation in the literature that used an RCT. Discuss how the RCT was conducted. Does the author discuss any limitations of using RCTs?

9. Why is the one-group pretest-posttest design a weak design according to the Campbellian validity typology? Why do community-based organizations frequently use this design nonetheless?
10. How do quasi-experimental designs differ from experimental designs? How are they similar?
11. Find an evaluation in the literature that used the nonequivalent comparison group design. Discuss how this design was applied. Did the authors mention the design as a weakness of their study?
12. Find an evaluation in the literature that applied the interrupted time-series design. Discuss how this design was applied and its strengths and weaknesses vis-à-vis internal validity and interpretation of the study's findings.

CHAPTER 11



The Holistic Effectuality Evaluation Approach to Outcome Evaluation

In spite of its popularity, the experimentation evaluation approach is not without criticisms. As discussed in the last chapter, the experimentation evaluation approach argues that randomized controlled trials (RCTs) or other experimental methods are the strongest design for outcome evaluation. However, many qualitative evaluators strongly disagree with this view. The next section will discuss the heated debates among evaluators over whether RCTs are the best method for program evaluation. The degree of controversy may indicate a need for an alternative outcome evaluation approach. This chapter will attempt to meet that need by introducing the holistic effectuality evaluation approach as both serving the theoretical foundations of the field and guiding real-world outcome evaluation. The chapter will also provide an empirical example to illustrate this approach's application and propose a rating system for judging the quality of a real-world evaluation. The relative strengths and limitations of the experimentation evaluation approach and the holistic effectuality approach will be discussed at the end.

ONGOING DEBATES OVER THE EXPERIMENTATION EVALUATION APPROACH

Proponents of the experimentation evaluation approach argue that RCTs are the gold standard for ensuring an evaluation's scientific rigor. The view has strong influence not

only on many evaluators but also on many decision makers. However, critics, especially qualitative evaluators, strongly oppose this view.

Evaluators' ongoing heated debates over whether RCTs are the best evaluation method can be traced back to the 1970s (Donaldson, Christie, & Mark, 2015). Many evaluators, often influenced by Campbell, view RCTs as the best method for providing rigorous evidence regarding the effectiveness of an intervention. They cite examples of false claims about interventions' effectiveness made by evaluations that used other methods. On the other hand, critics, especially from the qualitative camp, disagree. The criticisms are summarized below (Chen, Donaldson, & Mark, 2011):

Criticism 1: Internal validity should not be the top priority. Cronbach (1982) strongly disagreed with the idea that priority should be given to internal validity. He claimed that internal validity is "trivial, past-tense, and local" (p. 137). He argued instead that external validity should have priority because it is future oriented and addresses issues more interesting to decision makers. Cronbach viewed as profitable an evaluation that could draw stakeholders' attention to relevant factors and influence their decisions. In this view, evaluations need to allow extrapolation from the evaluation's specific populations, treatments, measures, and settings to others that are of interest to decision makers. For instance, an evaluation might include specific sites and clients, but a decision maker might need to determine whether to implement the program at another location with a different type of clients. A growing movement is encouraging more emphasis on external validity in evaluations (Chen, 2005; Green & Glasgow, 2006; Wandersman et al., 2008).

Criticism 2: RCTs are not the best method for obtaining credible evidence related to the kinds of questions evaluators should address. Critics often draw on one or more of the following:

- *Methodological criticisms.* For example, Scriven (2008) contended that because double-blinding is essential for quality RCTs but cannot be done in typical program evaluations, RCTs are susceptible to expectancy effects. In a double-blind experiment, neither participants nor researchers know whether a participant is receiving the treatment. Expectancy effects occur when knowledge of whom is receiving the treatment influences a participant's response or a researcher's measurement or interpretation of observed results.

- *Practical considerations.* For example, RCTs are often difficult to implement in the field, especially within the decision time frame given.
- *Contextual considerations.* For example, Greene (2009) joined Cronbach (1982) in criticizing narrowly defined evaluations that ignore the power of contextual influences on an intervention. Because they focus on variables that are under the experimenter's control, RCTs may overlook contextual factors.

Criticism 3: Persuasion, interpretation, and the subjective nature of conclusions—not validity—should be emphasized. Another general criticism of the Campbellian validity typology involves a fundamental question about the basis of validity. Some critics say that this view of validity is anchored in positivist and postpositivist thought. They instead embrace the phenomenological paradigm's viewpoint and promote naturalistic inquiry to understand inductively and holistically the human experience in natural settings (Guba & Lincoln, 1989; Maxwell, 1992).

In these debates, advocates of the experimentation evaluation approach have often defended RCTs on the basis of their rigor and scientific validity and challenged critics from the opposing camp to provide alternative methods that can provide equally credible evidence. The scientific defense of RCTs can be a difficult one for critics to counter.

Recently, arguments have been made to revise or expand the Campbellian validity typology to better serve evaluation, as discussed in volume 130 of the New Directions for Evaluation series, *Advancing Validity in Outcome Evaluation: Theory and Practice* (Chen, Donaldson, & Mark, 2011). In this volume, Gargani and Donaldson (2011) and Mark (2011) argued that the Campbellian validity typology is limited in terms of external validity, and they suggested strategies for enhancing external validity. Reichardt (2011) and Julnes (2011) offered ways to enhance precision by reclassifying the Campbellian typology. House (2011), Greene (2011), and Chen and Garbe (2011) argued that the scope of the Campbellian typology is too narrow to serve evaluation and proposed a revision or expansion of the typology to better serve evaluation. However, Shadish (2011), a collaborator of Campbell's who represented the typology's tradition in the volume, was not convinced by the others' arguments and stood against any change to the typology.

It is important to point out that the ongoing heated debates over RCTs and the experimentation evaluation approach take place mainly at the theoretical level. They have little effect on evaluation practice. Evaluation practitioners still follow the experimentation evaluation approach regardless of whether they agree or disagree with the approach when conducting outcome evaluations, because there has been no alternative outcome evaluation approach for them to adopt.

EFFICACY EVALUATION VERSUS EFFECTIVENESS EVALUATION

The points of debate outlined above can be better understood in the context of what kinds of evaluation evaluators are doing. Evaluation literature distinguishes two kinds of outcome evaluation: efficacy evaluation and effectiveness evaluation. An *efficacy evaluation* assesses the effect that an intervention has when conditions are ideal. The type of evaluation usually involves RCTs. According to Flay (1986) and Glasgow et al. (2003), efficacy evaluations are characterized by strong control, in that a standardized intervention is administered in an uniform way to a specific and narrowly defined homogeneous target group. RCTs are used to rule out threats to internal validity. Due to strong control and standardization, the evaluation can provide highly convincing evidence of the effect of an intervention.

An *effectiveness evaluation* assesses the effect of an intervention in real-world conditions. It is a nonrandomized study. According to Flay (1986), effectiveness evaluation is characterized as the standardization of availability and access among a broadly defined population while allowing implementation and level of participation to vary on the basis of real-world settings. This book uses these two terms, *effectiveness evaluation* and *real-world outcome evaluation*, interchangeably.

As discussed in the last chapter, the experimentation evaluation approach strongly favors efficacy evaluation because it is capable of ruling out threats to internal validity. The aim of efficacy evaluation is to manipulate conditions to ensure that a precisely measured “dosage” of the intervention is delivered, in standardized fashion, to clients. Tight research controls ensure the continued integrity of intervention conditions and study design throughout the evaluation. Thus, evidence from efficacy evaluation is usually very precise and highly defensible from the traditional scientific perspective.

Effectiveness evaluation has many drawbacks in terms of internal validity issues. To conduct an effectiveness evaluation is quite challenging because the real world is a messy place. As an effectiveness evaluation proceeds, a program’s usual implementers often are not as highly trained as those in efficacy evaluation. Clients are often not highly cooperative or highly motivated for change. External factors that are unrelated to an intervention and could confound evaluation results are often uncontrolled. Effectiveness evaluation is far less rigorous than efficacy evaluation. However, because effectiveness evaluation is carried out amid the messiness of the real world, it may have advantages in external validity. As will be discussed in Chapter 15, advocates of the experimentation evaluation approach, which holds internal validity as being of prime importance, recommend conducting effectiveness evaluation only after an intervention is assessed by efficacy evaluation.

This view puts evaluators in a very difficult situation. Evaluators are often asked to evaluate real-world programs. Since it is very difficult to do RCTs when evaluating real-world programs for ethical reasons and/or due to administrative or resource constraints, evaluations of an intervention's effectiveness usually apply a quasi-experimental or pre-experimental design. According to the experimentation evaluation approach, such real-world evaluations provide limited credible evidence. This view is further supported by the rise of the evidence-based intervention movement across many disciplines (Donaldson et al., 2008). Currently, advocates of the evidence-based movement claim that only those interventions evaluated by RCTs are evidence-based interventions or practices that are worthy of dissemination (Nutbeam, 1999; Speller, Learmonth, & Harrison, 1997; Stephenson & Imrie, 1998; Tilford, 2000). According to this school of thought, few evaluators' effectiveness evaluations of real-world programs are worthy of dissemination.

Most evaluators face the following dilemma when conducting outcome evaluation: On the one hand, evaluators have to apply the experimentation evaluation approach to effectiveness evaluation. On the other hand, the theoretical underpinnings of this approach hold that their effectiveness evaluations (real-world evaluations) are not of high quality. To address this dilemma, we must first reflectively examine the relationships among the experimentation evaluation approach, the Campbellian validity typology, and the nature of evaluation. This discussion will serve as a basis for introducing an alternative outcome evaluation approach that may alleviate evaluators' dilemma when practicing outcome evaluation.

RELATIONSHIPS AMONG THE EXPERIMENTATION EVALUATION APPROACH AND THE CAMPBELLIAN VALIDITY TYPOLOGY

As mentioned in Chapter 10, a close relationship exists between the experimentation evaluation approach and the Campbellian validity typology. The experimentation evaluation approach advocates an application of the typology for conceptualizing and guiding outcome evaluation. Because of the close relationship, any criticism of the experimentation evaluation approach can be easily interpreted as a criticism of the Campbellian validity typology. Perhaps a clear analysis of the distinction of these two may reduce this misunderstanding and provide clues to how to address the dilemma.

As mentioned in the last chapter, historically speaking, advocates of the experimentation evaluation approach helped introduce the Campbellian validity typology to evaluators. The typology provides evaluators with concepts,

principles, and methods for conducting outcome evaluation. However, it is important for evaluators to reflectively examine the following questions: Should evaluators view the Campbellian validity typology as the entire foundation for conceptualizing and guiding outcome evaluation?

To answer these questions, evaluators must first realize that the Campbellian validity typology, in spite of its usefulness for evaluation, was developed for “experimental research” rather than for evaluation purposes (Chen, 2010; Chen & Garbe, 2011). Campbell never formally established the typology for evaluation purposes. In fact, all three books of Campbellian validity typology (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002) hardly mention the word *evaluation*. Campbell, who was amazed to see the typology become widely popular in the evaluation community, became an overnight sensation in program evaluation. He was called an “accidental evaluator” (Shadish & Luellen, 2004). Today Campbell is regarded as a seminal figure in program evaluation (Alkin, 2013; Shadish et al., 1991).

All of us in the evaluation field appreciate the Campbellian validity typology’s contribution to evaluation, regardless of whether the contribution was intentional or unintentional. It is expected that this typology will continue to play an important role in evaluation. However, the above analysis may prompt evaluators to consider the following issues: If the Campbellian validity typology was not intentionally developed for evaluation purposes, are its relevance and contribution to evaluation purposes inherently limited? This is a legitimate question to ask. Because the typology was developed for experimental research, it may neglect issues important to evaluation. Another related question is this: If evaluators scrupulously follow the Campbellian typology, would this adherence lead evaluators to consider outcome evaluation issues only narrowly within the framework of experimental research? That is, evaluators may ignore issues that are trivial in terms of experimental research but crucial in terms of evaluation.

These questions are difficult for us to discuss or answer. As mentioned in the last chapter, most evaluators, including the author, have been trained under the influence of the experimentation evaluation approach. We are so accustomed to this frame of reference that it may prevent us from thinking about outcome evaluation issues creatively, blocking our ability to conceive of alternative approaches. The author believes that evaluators must think about outcome evaluation outside the proverbial box in order to better understand and practice real-world outcome evaluation.

The above discussion should not be interpreted as meaning that the author views the Campbellian validity typology as unsuited for evaluation. On the contrary, the author believes the typology is useful for many evaluations, including the alternative approach that will be discussed in the next section. However,

the author disagrees with the experimentation evaluation approach due to its rigid application of the entire principle, theory, and methodology of the Campbellian typology to outcome evaluation. This kind of application may be appropriate for research-oriented evaluation, but it is not appropriate for real-world evaluation. A dogmatic application of the typology may lead evaluators to compromise the nature and purposes of evaluation just to meet the requirement of applying the typology. Instead, the author argues we should critically consider which portions of the typology are fruitful or not fruitful for real-world evaluation and selectively and creatively apply the typology for evaluation purposes. We should put the typology to work in the service of evaluation rather than compromise our evaluations to apply the typology.

Furthermore, the above criticisms of the experimentation evaluation approach should not be taken as a prohibition against using this approach. On the contrary, the author believes the approach has its values and will continue to be used in evaluation. For example, there are situations when stakeholders do want evaluators to provide rigorous evidence on whether an intervention has affected outcomes. Under those circumstances, the experimentation evaluation approach is the one for evaluators to follow. The purpose of this chapter is to provide an alternative approach, called the holistic effectuality approach, for evaluating real-world programs. As will be discussed later, the holistic effectuality evaluation approach should not be viewed as universally better than the experimentation approach or vice versa. The last section of this chapter will discuss the relative strengths and limitations of these two approaches. Evaluators and stakeholders should be aware of these strengths and limitations and should select the approach that better serves their needs. It is in this spirit that the following holistic effectuality evaluation approach is presented.

THE HOLISTIC EFFECTUALITY EVALUATION APPROACH

The holistic effectuality approach falls under the integrated evaluation perspective discussed in Chapter 1. To recap, this perspective proposes that evaluators should synthetically integrate the dynamic nature of an intervention program in a community and stakeholders' views and practices with existing scientific methods to develop indigenous concepts, theories, and methodologies for program evaluation. This is a stakeholder-centered theory and methodology designed to achieve real-world outcome evaluation. Based upon this perspective, the *holistic effectuality evaluation approach* proposes a conceptualization of outcome evaluation that is very different from the experimentation evaluation approach's conceptualization. The differences between these two approaches are discussed as follows.

The Experimentation Evaluation Approach's Conceptualization of Outcome Evaluation

As discussed in Chapter 10, evaluators have traditionally followed the experimentation evaluation approach to conceptualize outcome evaluation. The traditional conceptualization of outcome evaluation includes the following features:

- Outcome evaluation is a conclusive evaluation; that is, it assesses whether an intervention brings about its intended outcomes.
- Outcome evaluation assesses an intervention's pure independent effects.

This conceptualization places great value on maximizing an evaluation's internal validity.

The Holistic Effectuality Approach's Conceptualization of Outcome Evaluation

The above conceptualization of outcome evaluation is a good fit for validity-focused outcome evaluation, as discussed in Chapter 10, in which evaluators' main concern is how internal validity issues are addressed or not addressed in an evaluation. The holistic effectuality evaluation approach's conceptualization of outcome evaluation is based upon lessons learned from past experience with evaluating real-world programs.

The holistic effectuality evaluation approach proposes the following conceptualization for real-world outcome evaluation:

- Real-world outcome evaluations are a hybrid type of evaluation that contains both constructive and conclusive assessments.
- Real-world outcome evaluation assesses an intervention's real-world effects, which are often neither pure nor independent.

These two premises are discussed in the following paragraphs.

Addressing the Issues of Real-World Evaluation Through Hybrid Evaluation

A real-world program, whether new or ongoing, is often not immediately suitable for outcome evaluation. It is impossible or unrealistic for evaluators to walk into a program and immediately conduct a conclusive assessment to

assess its merits. As discussed in Chapter 9, the SMART goals approach implies that a program may not have measurable goals. The evaluability assessment approach indicates that a program may not be ready for outcome evaluation. The plausibility assessment/consensus-building approach suggests that not all program goals are appropriate for evaluation purposes and that stakeholder groups may not agree on which goals should be evaluated.

Both stakeholders and evaluators have important reasons for a real-world outcome evaluation to be a hybrid evaluation. That is, evaluators first conduct a constructive evaluation to enhance a program's coherence so that they can later conduct a meaningful conclusive evaluation to assess its effectiveness. Evaluators cannot conduct meaningful evaluations if the program does not have measurable goals, the program lacks evaluability, or key stakeholders do not agree on which goals to evaluate. At the same time, it is challenging for stakeholders to put a program together so that it has a coherent structure. They see evaluators as having the expertise and tools that can help them make their program more coherent and thus more successful and more evaluable. Furthermore, stakeholders need to dedicate a lot of organizational resources and make a lot of organizational adjustments to help evaluators conduct an evaluation. Therefore, they believe that evaluators should contribute to program coherence as well as conduct an outcome evaluation. Stakeholders would be very uncomfortable if the evaluator behaved like a bystander—gathering pretest and posttest data, observing the success or failure of the program, submitting an assessment report, and finally just disappearing. Stakeholders would benefit little from such a purely conclusive outcome evaluation.

In conducting a real-world evaluation, evaluators need to discuss with stakeholders how the type of evaluation will transition from constructive to conclusive evaluation and what that means, as discussed in Chapter 1. My experience with evaluations indicates that stakeholders understand and support the need and reason for evaluators to transition from one evaluation type to another. Interestingly, stakeholders are usually more receptive to a hybrid evaluation than a conclusive evaluation, even when the conclusive assessment of a hybrid evaluation shows the program is ineffective.

Incorporating Adjuvants Into Real-World Evaluation

As discussed in Chapter 1, there are at least two major ways to think about how a social or behavioral intervention works in the real world. The first is that an intervention alone is sufficient to bring about changes. In this view, a behavioral or social intervention is like a dose of medicine or some other biomedical intervention. This line of thinking assumes that if an intervention is

delivered to clients, the intervention will make the desired changes. Advocates of the experimentation evaluation approach tend to hold to this line of thinking. The second way of thinking is that an intervention is necessary but not, by itself, sufficient for bringing about change. Behavioral and social interventions operate within a social system. Unlike the effectiveness of a biomedical intervention, the effectiveness of a behavioral or social intervention is likely to be conditioned by external factors such as culture, norms, social support, poverty, and so on. To address these external factors that are an inevitable part of the messiness of the real world, *adjuvants* are needed. The adjuvants motivate clients to work harder or make an intervention more appealing to clients. As discussed in Chapter 1, practitioners and other stakeholders usually hold this view, that a program operates in an impinging context and requires adjuvants.

Since stakeholders use adjuvants to make an intervention work, an evaluation designed to assess an intervention's pure independent effects may not be fruitful for a real-world program; it may be downright unrealistic. As will be demonstrated later, the adjuvants used by stakeholders often represent the many threats to internal validity that the experimentation evaluation approach so avidly wishes to eliminate. Thus, the experimentation evaluation approach can eliminate adjuvants from an evaluation by conceptualizing them as threats to internal validity, but because the adjuvants are essential to the execution of the intervention, such an evaluation may not be relevant to the stakeholders. Technically, an evaluation could assess an intervention's pure independent effects, but it may not accurately reflect the real world of the intervention and therefore not be useful to stakeholders.

An alternative approach is needed that uses the above conceptualization of real-world evaluation to assess the impact of an intervention realistically. The holistic effectuality evaluation approach, which includes constructive and conclusive assessments, is intended to serve this purpose.

CONSTRUCTIVE ASSESSMENT AND CONCLUSIVE ASSESSMENT: THEORY AND METHODOLOGY

The above discussions serve as a foundation for further development of the holistic effectuality evaluation's theory and methodology. As discussed in the last section, the holistic effectuality evaluation approach is a hybrid evaluation including both constructive and conclusive assessments. The constructive assessment is for the purpose of assessing a program's coherence and capacity to facilitate stakeholders in ensuring a program's coherence before conclusive assessment. The conclusive assessment then evaluates the joint effects of an

intervention and its adjuvants. The following text will begin by discussing the constructive assessment, which is relatively straightforward, and then explore the conclusive assessment, which requires more explanation.

Constructive Assessment

The constructive element of the holistic effectuality evaluation approach consists of the following components:

- It conducts a constructive outcome evaluation to ensure the coherence of a program, assess and build evaluation capacity, and conduct a participatory evaluation.
- It uses program theory or logic models to work with stakeholders to check whether the program has potential vulnerabilities.

Conducting a Constructive Outcome Evaluation to Ensure the Coherence of a Program

Evaluators need to conduct a constructive outcome evaluation, as discussed in Chapter 9, to assess whether a program has a coherent enough structure that a conclusive outcome evaluation can be effective. If a program does not have a set of clear and measurable goals, evaluators could use the SMART goals framework to facilitate stakeholders in developing them. If a program is not organized in the way it pursues program goals or is not evaluable, evaluators could apply the evaluability assessment to help stakeholders strengthen program structure. Similarly, stakeholders may be concerned about the plausibility of program goals or lack consensus about which goals should be evaluated; evaluators can apply the plausibility assessment/consensus-building approach to address these issues.

Assessing and Building Evaluation Capacity

In a real-world outcome evaluation, evaluators need to provide evidence regarding whether the implementing organization has evaluation capacity. If not, then evaluators need to facilitate stakeholders in building the program's evaluation capacity. There are two related reasons for including this element. The first is that stakeholders put a lot of effort into working with evaluators and do not want their efforts to go to waste. One of their expectations is that evaluators will enhance evaluation capacity. The second reason is that evaluators often need stakeholders'

help to collect the data needed for real-world evaluation. Without first building evaluation capacity, evaluators may not be able to collect accurate data.

Every organization collects a lot of data for administrative or managerial purposes. Evaluators cannot simply assume these data are valid for evaluation purposes. As an example, evaluating a media campaign's antismoking effect required hospitals in a community to share their records of identified smokers with evaluators. Evaluators assumed that these organizations, which were experienced with research studies, would not find such sharing difficult. But this assumption turned out to be incorrect. The evaluators received information, but they soon found that a large proportion of the individuals named in the hospital's records actually did not use tobacco. Setbacks such as unreliable data can be avoided by considering early on whether an implementing organization requires capacity building.

Program administrators and implementers play an important role in evaluation, so evaluators would be wise to ensure that an implementing organization as a whole has adequate understanding and skills to support evaluation activities and the capacity to deliver services as intended. If either is lacking, evaluators have tools to help organizations build such capacity. Evaluators and stakeholders need to discuss the following four specific evaluation capacities and decide which need to be assessed.

- *Capacity to enforce eligibility criteria and effectively recruit/screen participants.* Outcome evaluation can produce flawed results if people outside the target group enroll for services. To illustrate, imagine a flu immunization program whose implementing organization intended to screen applicants and admit only high-risk individuals, like those with asthma or heart disease. However, the screening procedure authorized by this implementing organization was unreliable, and the program accepted many participants from outside its target group. Unless this shortcoming is discovered before the intervention is delivered and data are collected, the integrity of the evaluation will suffer.

- *Capacity to record client and service data precisely.* Effectiveness evaluation requires implementers and service providers to precisely record their daily activities. In terms of service delivery, four questions are vital: Who is being served? What are clients' social and demographic characteristics? What services are provided? What is the level of client participation? It may be a mistake to assume that implementers will be able to answer these questions. Many evaluations require, for example, data concerning the race and ethnicity of each client. It is possible, however, that the implementers will provide very different numbers pertaining to race than to ethnicity, suggesting erroneous data. If this

happens, the evaluator needs to facilitate an effort by implementers to build their data-collecting capacity. Similarly, evaluators often rely on implementing organizations to store and manage client and service data. A wise evaluator will be certain this capacity will actually be there when needed.

- *Capacity to collect or assist in collecting outcome data.* Stakeholders who involve evaluators in program development from the beginning will likely enjoy significant benefits for identifying measurable goals. Evaluators or implementers may gather evaluation data pertaining to outcome measures; if implementers take on the responsibility, then the issues of organizational capacity discussed above are again applicable.

- *Capacity to use evaluation results for program improvement.* One major purpose of effectiveness evaluation is to generate information for use in improving current or future programs. The evaluator must ask, then, if the stakeholders seem capable of deploying evaluative information in a way that will benefit their program. If they do not, the evaluator should work to build their capacity to put evaluation data to good use.

Conducting a Participatory Evaluation

The purpose of this component is to gain stakeholders' input into the evaluation designs and methods that will be used for both constructive and conclusive assessment. Conducting a real-world evaluation as a participatory evaluation will ensure buy-in from stakeholders and results that are relevant and useful to stakeholders' practice. The holistic effectuality approach expects evaluators to engage with program managers, staff, and other stakeholders to understand their views and concerns and address them in an evaluation design.

In the absence of a participatory evaluation, stakeholders may mistrust evaluators. Stakeholders may wonder whether evaluators will fairly assess their program or whether evaluators have a hidden agenda. These suspicions could lead to withdrawing their support of or cooperation with evaluators. That eventually damages the quality and usefulness of a real-world outcome evaluation or even causes the entire evaluation effort to disintegrate.

An example is found in an assessment undertaken to determine who used a community park. Research staff showed up to the park daily and recorded the types of people there and the activities they participated in. However, no one from the park district or the evaluation team had communicated with community leaders or nearby residents about the purpose of the data-collecting effort. Park users consequently felt uncomfortable with the evaluators' daily

presence. It was rumored that the evaluation was, in fact, a move by authorities to manufacture some reason to close the park. Soon, a public outcry demanded the evaluation be stopped.

Another benefit of engaging with stakeholders is that evaluators can better understand the clients' culture, or subculture, and this knowledge may enhance the quality of the evaluation. Take an HIV-prevention program as an example. The goal of such programs is usually to increase safe-sex practices within a high-risk population, for example, Hispanic migrant workers. This goal is attractive and agreeable. The outcome measures used to evaluate its degree of success may, however, be more controversial. Literature tends to define safe sex as the use of a condom in *each* act of sexual intercourse by asking the following question: The last time you had sex, did you or your partner use a condom? However, for many program directors and implementers, a different interpretation of safe sex reigns, informed by their understanding of Hispanic culture. It is not realistic to expect high-risk individuals to use condoms for each act of intercourse without first specifying what kind of partner was involved. Among Hispanic migrant workers, for example, it can be deemed inappropriate, or even an insult, for a woman to ask a monogamous partner, such as her husband, to use a condom. So, an evaluation that arbitrarily measured safe sex in terms of harm *elimination* might be criticized by program stakeholders for overlooking the culture of the population being served. Fortunately, by engaging productively with stakeholders, evaluators can often find a solution acceptable to all. They can prompt the HIV-prevention program stakeholders, for example, to consider asking multiple questions when evaluating condom use: perhaps one about sex with the ostensibly monogamous partner, reflecting a harm-reduction orientation, and another about sex with casual partners, reflecting a harm-elimination orientation.

Evaluators also need to put to rest a myth about stakeholders—that they are not interested in the scientific credibility of an evaluation. For example, a national evaluation system was developed to follow CDC-funded, health department-based anti-HIV programs. During development, stakeholders were heavily involved in shaping the scope and focus of the evaluation system. They helped determine how evaluation questions should be phrased to reflect the reality of the populations they served, and they helped determine how evaluation results ultimately would be used. The stakeholders also agreed with evaluators on the value of using rigorous quantitative designs to assess programs' outcomes (Chen, 2001).

Evaluators have to spend more time and effort to do a participatory evaluation than a nonparticipatory one, but the participatory approach tends to have the following benefits:

- It helps evaluators better conceptualize the program.
- It makes possible the precise definition of the intervention and goals (during intervention planning).
- It determines a context for the use of evaluation findings.
- It prevents the substitution of evaluator preferences and values for those of local stakeholders.
- It ensures the collection of information relevant to a program and useful to stakeholders.

Using Program Theory or Logic Models to Check for Potential Vulnerabilities

To plan and deliver an intervention program in the real world is highly challenging. Too many factors inside or outside a program could interpose difficulties between the program and its intended outcomes. Program designers will likely overlook some vital elements or fail to take precautions against certain risks.

One important strategy evaluators can use is to walk through the action model/change model schema or logic model discussed in Chapter 3 and brainstorm with stakeholders which components, elements, and/or linkages are vulnerable. With this useful information, stakeholders can then determine whether to take remedial action. Stakeholders expect evaluators to use evaluation tools to strengthen their programs and are grateful for evaluators' help. Such assistance also helps to build a collaborative relationship between stakeholders and evaluators. An outreach program to serve sex workers offers an example. A program manager in charge of the project overlooked an ongoing police campaign to clear several neighborhoods of sex workers. The manager directed staff members to conduct outreach among sex workers in these neighborhoods and the rest of the community. A number of the unsuspecting staff members were actually arrested, and a great deal of effort was needed to convince police of their innocence. In this case, if evaluators had used program theory or logic models to work with stakeholders to check for potential vulnerabilities, they would likely have identified this shortcoming, that the program was not notifying police departments of its target population and service. Clearly, evaluators should help stakeholders be proactive and avoid potential problems.

Similarly, an education program, for example, was built around an innovative new textbook with a tight publication schedule. When evaluators brainstormed with stakeholders which components, elements, and/or linkages were important to a program theory or logic model, it would be very reasonable for the evaluators to raise the question of what would happen if the textbooks were not available on time. Program management and staff are usually grateful for any

concerns raised by evaluators. Moreover, evaluators' willingness to help explore potential vulnerabilities and discuss them with stakeholders can improve communication and even enhance the trust between evaluators and stakeholders.

Conclusive Assessment

The holistic effectuality evaluation approach proposes a conclusive assessment that is very different from that of the experimentation evaluation approach, especially in terms of how it addresses threats to internal validity. The holistic effectuality evaluation approach argues that a real-world program has the following features:

- Stakeholders often use adjuvants to make an intervention work.
- These adjuvants usually constitute threats to internal validity, as discussed in Chapter 10.

These ideas were inspired by the author's experience of attending a Zumba weight loss program. To help readers understand the above arguments, the following section provides a detailed description of the author's experience and lessons learned.

THE ZUMBA WEIGHT LOSS PROJECT

I was overweight. Given the overwhelming evidence that obesity causes illness, I decided a couple of years ago to lose some pounds. I signed up for a Zumba weight loss program offered by a fitness center partly because of my fondness for rhythm. After enrolling in the program, I suddenly remembered what had happened in a meeting with community practitioners a few months ago. During the meeting, a few practitioners had commented that outcome evaluation is too academic and not relevant to what they were doing. I was perplexed by their comments, and despite intensive discussions during and even after the meetings, I still was not sure what they meant. I believed part of the reason for my lack of understanding might be that I lacked firsthand experience as a practitioner or client in an intervention program. That situation could now be changed. By joining the Zumba program,

(Continued)

(Continued)

I became a client of an intervention program. I felt that participating in the Zumba program would provide me with a good chance of understanding the dynamic relationship between a real-world program and evaluation.

I discussed with the instructor my idea for doing a participant observation project in the Zumba classes. I would have a dual role in the classes: attendee and researcher. My research purpose was to learn about program implementation and effectiveness from the standpoint of an instructor and a client. I explained to the instructor that, as a participant-observer, I would conduct my research activities relatively unobtrusively by observing class activities and interaction patterns among the instructor and participants. I would occasionally ask her some questions, but I would not interview classmates. The instructor agreed with the idea under the condition that I would not interrupt the classes and would not bother other attendees. I happily agreed. I took notes on my observations and my experience during each class only after I went home.

Each Zumba class consisted of 1 hour of intense exercises using choreography inspired by Latin, hip-hop, Bollywood, and other international music. The program included weight checks, Zumba dance classes, and the instructor's occasional feedback regarding attendees' weight-loss progress. The instructor wanted attendees to come to class 3 times per week.

Before the program, I conducted a mental exercise to imagine how evaluators would apply RCT to assess the effect of a Zumba program on weight loss. Evaluators would recruit a group of homogenous participants who expressed enthusiasm about attending Zumba classes. The participants would be compensated for their time. They would be randomly assigned to an experimental and a control group, with Zumba classes offered only to the experimental group. Class instructors would be highly trained and skillful, as rated by previous students and supervisors. The instructors would exactly follow the protocol of Zumba classes, and each class would be supervised to ensure this was the case. All participants in the experimental group and the control group would be weighed before attending the program and 6 and 12 months after the program's conclusion. Through random assignment to treatment and control groups and measurement before and after the treatment, threats to internal validity would be minimized. The evaluation would provide rigorous evidence about the Zumba program's pure independent effects on participants' weights.

The above mental exercise provided me a reference point for my participant observation of the Zumba class in the real world.

After the program started, the instructor asked participants to weigh themselves and record their weights before each class. I knew that according to the experimentation evaluation approach, this would be a “testing” threat to internal validity. It could confound the evaluation of the intervention’s effect. However, regardless of whether it is a threat or not, this is an ingredient of real-world practice for any weight-loss class. In other words, like it or not, “testing” is part of the real-world implementation of a weight-loss program. RCTs can eliminate the threat from testing, but from the instructor’s and participants’ point of view, testing is an essential part of the program. This was just the beginning of my journey through a real-world program.

The class was predominantly female, with about 70 females and only 3 males, including me. On the first day, the instructor chatted with the three males before the class. She asked us if we had joined the class because our wives’ had pushed us to do so. My two male classmates said yes. No wonder their faces wore unhappy, stony expressions. I was the only male who said no. I was puzzled why she asked this question, so I chatted with her after the class. She said these two male students would quit the class soon because they lacked motivation. She was correct. After the first week, they quit the program. I was the only male remaining in the class.

Again, as an evaluator, I recognized this as *self-selection bias*, a threat to the internal validity of an evaluation. People who are likely to be successful tend to participate in, or complete, an intervention. When this participation is not controlled in some way, it can confound evaluation findings. According to the experimentation evaluation approach, evaluators must rule out selection bias. However, my observation showed me that this threat is also an integral part of real-world practice. The instructor wanted the three males to persevere in the class. However, because there were so many participants, she only had a limited amount of time to work with any one participant. Having an idea of who was likely to stick with the program helped her allocate her time for maximum effect. No wonder, then, that she appeared to pay more attention to me than to the other males. Another interesting thing she mentioned was that it was very difficult for her to work with participants who are not committed to change. She commented that they needed help that was beyond her capacity to give. In other words,

(Continued)

(Continued)

participants with low motivation would need a great deal of attention and resources devoted to tailor-made intervention. A program delivering these features would be very expensive. Her comment implied that self-selection reflects an important truth about real-world programs: Practitioners must view self-selection as a typical problem that they must address in practice rather than eliminate or ignore.

Furthermore, I observed that self-selection was not a one-time issue, occurring only at the beginning of the program. Rather, it was a recurrent issue that the instructor had to deal with in every class. Participants who had higher motivation, or believed in the class, tended to take places close to the front of the room, while those who were less motivated tended to be in the last row, as far as possible from the instructor. As a result, those in the front rows got a good view of the instructor's Zumba steps, quickly learned the steps, and enjoyed the dances. Those in the back of the room did not see the instructor's demonstration clearly, lagged behind in learning the steps, and were frustrated as they tried to dance. As a consequence, they were more likely to quit. The instructor was obviously aware of the problem, as she invited those participants in the last rows to the front to help stimulate their interest. In other words, the instructor used adjuvants to address self-selection issues in running the class. By observing self-selection as it occurs in a real-world program, I began to wonder about the wisdom of eliminating self-selection from the evaluation scenario because of the threat it poses to internal validity, as suggested in the evaluation literature. Technically, evaluators could rule out this threat by using RCTs, but in consequence, the structure of the activity being evaluated would not resemble real-world practice. Perhaps the experimentation evaluation approach screens out important information that is useful to stakeholders.

After I had attended the class for 1 month, my body weight remained the same. I expressed my frustration to the instructor. She asked me to weigh my body after a class session and especially after taking a shower. Again, I immediately recognized this posed the *instrumentation* threat to internal validity; the switch from the preclass measure of body weight to the postclass measure of body weight could bias the evaluation. However, as a class attendee, I decided to give it a try. The after-class measure showed I was about 1.5 pounds lighter, perhaps due to perspiration. The instructor said this indicated that my body had good potential to lose 1.5 pounds, if

I kept up the exercise. Interestingly, she used the change of measurement tool, with its instrumentation validity threat, to motivate me to continue the class. It worked. After another 2 months, I found I had lost 3.5 pounds according to the postclass measure. Eventually, I switched back to weighing myself before class; the scale still showed a 2-pound weight loss. I was amazed that the instructor had creatively used the invalid measurements of differing instrumentation as an adjuvant to motivate me. This case indicates that practitioners can use a knowledge of threats to internal validity to benefit their clients.

I reported my progress to the instructor. She was happy, too, and commented that I would soon see even more progress. The program had started just before the holiday season, including Thanksgiving, Christmas, and New Year, when people tend to eat and drink more than usual. If I continued to exercise, she predicted I would lose a few more pounds starting in the new year. Her prediction was correct again. I realized this phenomenon as the threat to validity of *maturation*, that is, a change due to a natural change in circumstances rather than an intervention. The instructor creatively used the maturation threat to validity as an adjuvant to encourage me. In fact, I lost another 2 pounds in January. At that point, I had firsthand experience with how important many threats to internal validity are to the success of practitioners and clients and wondered even more whether application of the traditional experimentation evaluation approach to a real-world program can overlook a great deal of information stakeholders would find valuable.

Meanwhile, I also observed how my instructor used some threats to internal validity as adjuvants to motivate other participants and run the classes. I noticed she built good, informal relationships with many of the female participants. They went out together a lot. These friendships helped the instructor build a fan base of participants in the class. This constituted a *history* threat to internal validity, according to the experimentation evaluation approach, but the instructor explained that this practice was a typical activity for her and other instructors she knew. Other events that posed a history threat also occurred. After observing my loss of 8 pounds in 6 months, the instructor suggested that I incorporate high-impact steps into my Zumba routines by taking broader strides and jumping higher. Also, because the program had been working for me, I decided to increase the number of classes I attended from 3 to 6 per week. These changes in the

(Continued)

(Continued)

intervention could pose fidelity or history threats to the validity of an evaluation, as traditionally viewed, but I felt these changes were essential to my continued success. In fact, I lost 12 pounds in 1 year and felt much healthier than before.

At that time, both the instructor and I felt I could build on my success so far by eating healthier foods. In addition to attending Zumba classes, I began to eat a lot of fruits, vegetables, and other healthy foods every day. Again, this action posed a history threat, according to the experimentation evaluation approach. Between the Zumba classes and all the adjuvants, I lost a total of 20 pounds (about 14% of my original body weight) in 18 months, without taking any medicine or undergoing surgery. I feel I am healthier and have more energy now than before I took the Zumba class. My friends and colleagues have noticed positive changes in my appearance and congratulated me on my success. Their encouragement provides me with extra motivation to continue to take Zumba classes and eat healthy foods. The success has been sustained. I have been able to maintain the 20-pound weight loss for over a year and am still going strong. I feel proud of the accomplishment.

Lessons I learned from the weight-loss program included the fact that both my instructor and I benefited from adjuvants (threats to internal validity). Let's recall the mental exercise in which I imagined using RCT to evaluate the efficacy of the Zumba program. In an RCT, all the efforts of the instructor and me to make the intervention successful would have been threats to internal validity, or confounding factors. All would have to be excluded from the evaluation. The evaluation that used RCT would show whether the Zumba program had a pure independent effect on weight loss—it would be rigorous—but the evaluation would hardly be relevant to the Zumba program in which I participated. As a matter of fact, it would provide little insight into how the program operates in the real world. The evaluation and its results would be very artificial. Stakeholders in different settings with different clients who were interested in adopting the program would learn very little from the evaluation.

At this point, I felt I had a better understanding of the comments made by the practitioners in the meeting that outcome evaluation (under the experimentation evaluation approach) is too academic and not relevant to real-world practice. Furthermore, I was convinced that evaluators must have an alternative approach to represent stakeholders' views and reflect their interest in real-world outcome evaluation.

WHY ADJUVANTS ARE NEEDED FOR REAL-WORLD PROGRAMS

Adjuvants are formally defined as factors used by stakeholders to enable an intervention to work or enhance its effectiveness in the real world. This section will discuss why adjuvants are important for contextualizing a social or behavioral intervention in the real world.

Intervention programs aim to change clients' behaviors and make their lives better. In doing so, these programs ask clients to give up habits or things they have enjoyed for many years and adopt new behaviors or habits that are physically or mentally challenging. In the Zumba project, I confronted many hurdles. Firstly, the hour of intensive exercise in each class was a physical challenge to me and perhaps many others, at least for many months. In addition, class attendance is time-consuming. It generally takes me nearly 3 hours to travel to class, exercise, change clothes, take a shower, etc. This means I have to set aside 3 hours a day for a Zumba class 6 days a week. In order to do so, I have had to give up many social gatherings and leisure activities, such as going to the theater and watching TV. Furthermore, I have to deal with my female classmates' teasing. For most of the classes, I am the only male there. My female classmates seem to enjoy needing me. For example, every time the instructor plays Beyoncé Knowles's "Single Ladies (Put a Ring on It)," my female classmates all point their fingers at me.

The challenges I face in the Zumba program are not unique to me. Most of health promotion programs and social betterment programs require clients to give up pleasures or simple habits and make tough behavioral changes. For example, an education program may ask students to give up time playing video games to study more instead. A nutrition program may ask fast-food lovers to give up delicious French fries and juicy hamburgers to eat foods such as broccoli and tomatoes. Even worse, clients are not asked to make these difficult changes just for one day but over the long haul.

Facing these challenges, practitioners and clients feel they need all the help they can get to make an intervention work. An intervention alone is less likely to be fruitful if it does not get extra help from adjuvants (threats to internal validity). These assists could include building trusting relationships, motivating clients, providing social supports, and even adding extra interventions. As shown in the Zumba project, these extra supports, which posed threats to internal validity, were essential for the instructor to run the class and for my success. In a real-world setting, many threats to internal validity are integral parts of intervention programs.

Making the situation more complex is the fact that clients often live in an environment that inhibits implementation of an intervention. For example, children in an educational program for disadvantaged families may live with parents with inadequate parenting skills, in a neighborhood where they encounter peer pressure

to join gangs or sell drugs, and in an area with few libraries or bookstores. Implementers who expect success cannot focus only on the intervention protocols and not use adjuvants to address some of these problems.

Types of Adjuvants and Threats to Internal Validity

Stakeholders use different types of adjuvants for different purposes. These types and their relationships with internal validity are discussed as follows.

Helping Clients Realize or Admit They Have a Problem and Need Help

Practitioners believe it is impossible to work with clients unless the clients realize or admit they have a problem and need help. One approach is to pretest clients about the problem. The pretest is not just to evaluate the client but also a way to present information that may convince the client that he or she has problems and needs help. For example, in the Zumba project, the instructor asked each participant to weigh themselves on a scale before the program started and used that measurement to discuss with participants the problems they might experience due to being overweight. This kind of adjuvant poses a testing threat.

Building Relationships With Clients

Practitioners may believe a trusting relationship with clients is essential for delivering services. Building relationships with clients can also motivate clients to continue participating in the program. In the Zumba example, the instructor built friendships with attendees, and these friendships motivated many participants to come to class often. This kind of adjuvant poses a history threat.

Enhancing or Supplementing the Intervention

Practitioners may feel an intervention needs supplementation to make it work. For example, instructors of an educational program may add a parental-involvement component to make the intervention work better. This kind of adjuvant can be used to address environmental issues. For example, in the Zumba program, the instructor encouraged participants to eat healthy foods, because continuing to eat unhealthy foods could wipe out clients' achievement in the Zumba classes. This kind of adjuvant poses a history threat according to the experimentation evaluation approach, but it is very useful for a real-world program.

Addressing Cultural Appropriateness Issues

An intervention may not be culturally competent with respect to the particular population it seeks to serve. Program directors and staff may need to modify the intervention or add adjuvants to make it culturally appropriate. This kind of adjuvant poses a history threat.

Encouraging Participation or Retention

In the real world, self-selection is an unavoidable part of the intervention process. People who want to make a change are more likely to participate in an intervention program. Policy makers, program directors, and staff need to decide what to do with this selection tendency. They can decide to use adjuvants such as admission criteria that screen out those who are less likely to have success or they can have a policy of first come, first served. Self-selection continues to operate after participants join an intervention program. Those who are more likely to succeed are again more likely to stay with the program; those unlikely to succeed tend to quit. Decision makers need to decide how to use limited resources to address retention issues. Depending on overarching values or the resources available, decision makers or practitioners may decide to use adjuvants to encourage those participants less likely to succeed to stay in the program or to make sure those who are likely to succeed continue with the intervention. In the Zumba program, the instructor spent more effort on ensuring those likely to succeed to stay with the program. This kind of adjuvant poses a selection threat.

Taking Advantage of Environmental Change or Natural Trends

The environment of an intervention program does not stand still. An event or natural trend in the environment may help or hinder clients' progress. Practitioners may recognize such developments and turn them to clients' favor. For example, in the Zumba program, the instructor recognized that people eat more during the holiday season and used that fact as an adjuvant to motivate me to continue the program and achieve a favorable result. This kind of adjuvant relates to the maturation or history threat.

It is important to note that adjuvants do not have to be positive. Sometimes, counselors or clients might use controversial or questionable adjuvants in a program. For example, the author visited a reputable alcohol abuse treatment center and found during a break that a large number of clients rushed into a break room smoking cigarettes heavily, which was allowed at the center, perhaps to relieve their craving for alcohol. To have a balanced evaluation, evaluators should assess whether this kind of adjuvant is used in a program.

METHODOLOGY FOR REAL-WORLD OUTCOME EVALUATION

The holistic effectuality evaluation approach proposes that the basic purposes of real-world outcome evaluation are to assess the joint effects of an intervention and its adjuvants and eliminate potential bias. This means that real-world outcome evaluations do not have to eliminate all the threats to internal validity, as is important in the experimentation evaluation approach.

Assessing the Joint Effects of an Intervention and Its Adjuvants

As discussed previously, outcome evaluation has traditionally been conceptualized by the experimentation evaluation approach as an assessment of an intervention's pure independent effects. The holistic effectuality evaluation approach proposes that the purpose of real-world outcome evaluation is to assess an intervention program's joint effects, that is, the effects of an intervention in combination with adjuvants, or mesological intervention. Joint effects are those that clients get in the real world, because programs typically use both an intervention and supplemental efforts. For example, a real-world evaluation of my participation in the Zumba class would find out I lost 20 pounds as a joint effect of the Zumba class and adjuvants, including my instructor's encouragement, my commitment to weight loss, a shift from low- to high-impact choreography, an increase in the number of classes, and a change in my diet.

A holistic effectuality evaluation has its strengths and limitations. One of the strengths is that the real-world program studied operates in a real community and with clients who experience real benefits. This approach provides an opportunity for evaluators to examine, learn, and report dynamic relationships between intervention and adjuvants in a real-world setting. This kind of information is useful for further advancing the science of intervention. Furthermore, since stakeholders are likely interested to know what clients are actually getting out of the program as a whole, they are likely to view the evaluation as highly relevant to their work. It is also possible that real-world effects are more likely than the effects of a tightly controlled experiment to be disseminated to other communities.

However, an evaluation of a real-world effect usually cannot answer the question "What was the contribution made by the intervention alone?" Researchers are particularly interested in knowing the precise relationship between an intervention and outcome for purposes of scientific advancement. An evaluation of a real-world joint effect does not address this interest.

For example, a real-world evaluation of my participation in the Zumba class would find out I lost 20 pounds. However, the weight loss was a joint effect. The real-world effect is very important to me and my instructor. Furthermore, I can definitely say that the Zumba class is the driving force behind my 20-pound weight loss. However, from a researcher's viewpoint, my opinion is not precise enough; a researcher would prefer an evaluation that could estimate how many pounds of weight loss are due to the Zumba class alone. I am not able to provide information about such a pure independent effect. If evaluators and stakeholders are interested in assessing an intervention's pure, independent effect, they should follow the experimentation evaluation approach, discussed in the last chapter.

However, an intervention's pure, independent effect may not reveal to practitioners what actually works. The strengths and limitations of evaluating an intervention's real-world effects versus its pure independent effect will be discussed later. However, it is important to realize that social and behavioral intervention programs require knowledge of both pure independent effects and real-world joint effects. The long-held myth that the evaluation of pure independent effects has a higher priority than evaluation of real-world joint effects is counterproductive to the development of intervention science.

Eliminating Potential Biases

The holistic effectuality evaluation cannot rely on the traditional methodological thinking used to support the assessment of pure independent effects. Real-world evaluation must be based on its own methodological principles.

- *Research designs that use manipulation and control devices are not appropriate for assessing real-world effects.* Thus, RCTs are not appropriate for assessing real-world joint effects. The manipulation and control used in RCTs destroy the natural texture of real-world effects.
- *Real-world evaluation does not require eliminating all threats to internal validity.* As discussed previously, if the factors that may threaten internal validity are used as adjuvants in a program, evaluators do not need to control for these threats.
- *Real-world evaluation still needs to eliminate potential biases.* Biases are those threats to internal validity that do not arise from adjuvants and that potentially could weaken evidence of program effectiveness. Generally speaking, the following threats to internal validity are likely to bias real-world evaluations: instrumentation, maturation, regression toward the mean, and attrition.

These three principles provide a new insight into which research designs to use in real-world evaluation. For example, RCTs are not the best design for holistic effectuality evaluation. Instead, quasi-experimental designs are strong designs for dealing with bias. For example, the experimentation evaluation approach indicates that the nonequivalent control group design suffers from a self-selection threat. However, if stakeholders formally use self-selection as an adjuvant in their program, the use of the nonequivalent control group design in evaluating the program would not be vulnerable to this threat. In addition, we may need to reassess the traditional view that pre-experimental designs are the weakest designs for outcome evaluation. For example, if real-world evaluation does not have the burden of eliminating all threats to internal validity, the one-group, pretest-posttest design, which provides baseline data for assessing change, can be considered an effective design.

RESEARCH STEPS FOR ASSESSING REAL-WORLD EFFECTS

The holistic effectuality evaluation approach proposes a set of research steps for assessing an intervention program's real-world effects. The research design consists of three major and one optional components:

- Inquiring into the process of contextualizing an intervention in a real-world setting
- Using a relatively unobtrusive quantitative design to address biases and assess change
- Using an auxiliary design to triangulate evidence
- Replicating a mesological intervention (optional)

Inquiring Into the Process of Contextualizing an Intervention in the Real-World Setting

The core of an intervention program is an intervention. However, as illustrated in the Zumba project, an intervention alone is not sufficient to bring about success. Practitioners and other stakeholders often contextualize an intervention through the use of adjuvants in order to make it work in the real world. Evaluators need to conduct an inquiry into whether stakeholders use adjuvants to make an intervention work. If so, what are these adjuvants, and what are their functions? In the inquiry, evaluators usually use qualitative methods, discussed below, to gather the following information: Are adjuvants

used? Why are they used? What are the relationships between these adjuvants and the intervention? What are implementers' and clients' experiences with these adjuvants?

In addition to looking for adjuvants, evaluators can also do an environmental scan to determine whether there are significant environmental events that may affect outcomes. Evaluators use this inquiry to form a model of a mesological intervention—that is, a combination of an intervention, adjuvants, and significant environmental events—for impact assessment. In the Zumba project, the mesological intervention included the following items:

- *Intervention.* Zumba classes
- *Adjuvants.* The instructor's positive reinforcement of participants' behaviors, the instructor's relationship with participants, the strategies used to deal with retention, and encouragement to eat healthy foods
- *Environmental events.* Seasonality (e.g., the holiday season)

Evaluators can apply one or more of the following qualitative methods (Creswell, 2013) to collect information on the mesological intervention:

In-Depth Interview

This method involves asking participants questions, listening to their responses, and posing additional questions to gain clarification or expanded understanding of particular issues. Questions are typically open-ended. Participants are encouraged to freely express their own views, perceptions, and opinions.

Focus Group Meeting

A focus group meeting can be viewed as an interview of a small group. Group size often ranges from 6 to 12 persons. Participants in a focus group are asked to provide their own opinions and comments, listen to what other members of the group have to say, and react to those members' views and observations. The major purpose of a focus group meeting is to elicit ideas, experiences, and insights in a group setting where participants stimulate each other and consider their own views along with the views of others. Usually, these interviews will be conducted several times, with different groups, so that interviewers can identify patterns in participants' opinions and perceptions. During a meeting, interviewers act as facilitators by introducing the subject, guiding the discussion, cross-checking each comment, and ensuring that all members have an opportunity to express their opinions.

Observational Methods

The purpose of observational methods is to allow evaluators to obtain thorough descriptions of program activities, interactions between participants and implementers, and the meanings participants and implementers have attached to the intervention program. These methods also give evaluators a better understanding of the context in which program activities happen. There are two popular observational methods: direct observation and participant observation.

- *Direct observation.* This method usually uses detached observers to systematically observe program activities, ideally in unobtrusive ways.
- *Participant observation.* This method allows observers to become members of the group or community being studied. Observers participate in activities and observe how people behave and interact with each other. By doing so, observers are able to see what is happening and feel what it is like to be a part of the community.

Using a Relatively Unobtrusive Quantitative Design to Address Biases and Assess Change

Based upon the discussion of the methodology for assessing joint effects of mesological interventions, the following are examples of research designs that are highly useful for a real-world evaluation.

One-Group Pretest-Posttest Design

As discussed previously, one-group pretest-posttest design is regarded as an adequate design for assessing joint effects. One of the merits of the design is that it provides baseline data for quantitatively assessing changes. In addition, it is relatively inexpensive and easy for community-based organizations to administer compared to quasi-experimental designs. Furthermore, as will be discussed in the next section, the holistic effectuality evaluation approach proposes an auxiliary design component to strengthen a quantitative design. Therefore, potential biases that cannot be addressed by the one-group pretest-posttest design can be alleviated by the auxiliary design.

Nonequivalent Comparison Design

The nonequivalent comparison design is used to add a comparison group to the one-group pretest-posttest design. Ideally, the comparison group will be

as similar to the intervention group as possible. This design is a stronger design than the one-group pretest-posttest design because it attempts to address the maturation problem. The design does not disrupt the organizational routine, and its logic is easy for stakeholders to understand. However, this design is much more expensive and resource-intensive than the one-group pretest-posttest design.

Interrupted Time-Series Design

The interrupted time-series design is another good design for real-world effectuality evaluation. This design can effectively control many biases. It requires conducting monthly, quarterly, or yearly outcome measurements before and after an intervention. Many programs may not have the data necessary for the application of this design.

Using an Auxiliary Design to Triangulate Evidence

Evaluators of real-world programs should creatively use additional research methods to collect additional qualitative and quantitative data, which can be used to triangulate the findings from the main quantitative design. The purpose of the auxiliary design is to strengthen the evidence found through the main design. This component usually uses the following strategies:

Qualitative Evaluation

Qualitative data related to implementers' and clients' experience with the intervention provide rich information for triangulating quantitative findings. If the qualitative data are consistent with quantitative findings, they provide stronger support for the evaluation's conclusions. If they are inconsistent, evaluators need to collect additional information to reconcile the differences.

Pattern Matching

Pattern matching is a detailed elaboration of theoretical patterns between the intervention and its outcomes for the purposes of empirical testing. This strategy was developed as part of the Campbellian validity typology. When patterns exist, the validity of any conclusions about the program's effect is increased. Pattern matching is analogous to comparing fingerprints. As argued by Trochim and Cook (1992) and Mark, Hofmann, and Reichardt (1992), the

more detailed patterns are, the easier it is to compare theoretical outcome data to observed outcome data—and the more evaluators' confidence is increased in their assessment of the program's effectiveness.

Pattern matching is particularly useful when a control or comparison group for an outcome evaluation is unavailable. A challenge, however, must be met: An outcome variable must be identified that (a) is not targeted by the intervention and (b) is subject to the same disturbances as the target outcome variable. If the changes are observed in both the target and nontarget outcome variables, then in theory, the disturbances rather than the intervention affected both. If the changes are observed only in the target outcome variable, then they are unlikely to have resulted from variables other than the intervention and adjuvants. Trochim (1984) illustrated pattern matching with the following example. If an algebra tutorial program for disadvantaged students has an effect, then in theory, the students' algebra scores should rise, but their scores in other subject areas, such as geometry, should not rise—at least not in the short term. Matching the score/subject pattern for algebra to score/subject patterns for other subjects should help solidify causal inferences about the effect of the algebra tutoring.

The rationale for pattern matching is the theoretical principle that all related outcome variables should experience identical influences from any confounding factors or disturbances. Thus, a student's scores for various mathematical subjects all should be identically influenced by his or her maturation, motivation for achievement, and experience of events (e.g., decreased parental involvement). Assuming the similarity of these influences, a pattern marked by a sudden increase only in the score targeted by intervention, with no increase in other scores, tends to confirm the intervention's effect.

Pattern matching can be combined with the one-group pretest-posttest design or other designs to strengthen evidence of program effectiveness. For example, McKillip (1992) combined pattern matching with an interrupted time-series design to evaluate a health promotion campaign. The purpose of the campaign was to encourage responsible attitudes and behaviors concerning alcohol in a university setting. Time-series data on three outcomes—responsible alcohol use, good nutrition, and stress reduction—were collected before and after the intervention. The targeted outcome variable was responsible alcohol use, so the nontargeted outcome variables were good nutrition and stress reduction. Theoretically speaking, if the campaign was exhibiting an effect, measures for responsible alcohol use should change, whereas those for good nutrition and stress reduction should not. The study found that observed evaluation patterns in fact matched the theoretical patterns very well.

Assessing the Underlying Causal Mechanisms

Potential users, or adopters, will feel more confidence in the outcomes of the program if the evaluation can illustrate how the causal mechanisms underlying the intervention operate. How to expand an effectiveness evaluation to assess underlying causal mechanisms is a topic of theory-driven outcome evaluation, discussed in detail in Chapter 12.

Replication of the Mesological Intervention (Optional)

A real-world outcome evaluation is much less expensive to conduct than RCTs or other experimental methods. If possible, evaluators are encouraged to replicate the mesological intervention (a combination of the intervention and adjuvants) at another site or multiple sites. The purpose of replication is not just to enhance external validity but to achieve a measure of internal validity. If the mesological intervention is found to be effective at one or more other sites, this finding strengthens the evidence for claiming an intervention program's effectiveness.

EXAMPLE OF A REAL-WORLD OUTCOME EVALUATION

To reduce the problem of HIV infection in China, the government implemented voluntary counseling and testing (VCT), conducted at local government health agencies in major cities in 2003. However, such government VCT sites had difficulty attracting high-risk populations, such as injecting drug users (IDUs) or sex workers, due to the fear of stigmatization or arrest for their illicit activities. Alternatives to government VCT sites were needed to serve IDUs and other high-risk populations. Unfortunately, there were very few Western style nongovernmental organizations (NGOs) in China to fill the gap. This innovative study (Chen et al. 2007) used a local Women's Federation, a government-organized NGO, to provide VCT services to IDUs.

The project used the holistic effectuality evaluation to design and conduct an evaluation of this real-world program. Following the hybrid approach, the evaluation started with a constructive assessment and then moved on to a conclusive assessment.

(Continued)

(Continued)

Constructive Assessment

Conducting Constructive Outcome Evaluation

As mentioned previously, during the research project, provincial health departments and the Chinese Center for Disease Control and Prevention were responsible for HIV prevention. The local Women's Federation had these two government agencies' approval to run the project. However, the evaluation team and the Women's Federation still felt it was crucial to apply the plausibility/consensus-building approach to make sure they agreed on the program and the goals to be evaluated. After two facilitated meetings, representatives of the Health Department and Center for Disease Control and Prevention not only agreed on the program and measurable goals for evaluation but also offered to train Women's Federation staff on VCT.

Using a Participatory Approach for Designing Outcome Evaluation

The evaluation team took a participatory approach to designing the conclusive outcome evaluation by inviting input from the Women's Federation, provincial health departments, and Center for Disease Prevention and Control. All parties agreed on the research design, outcome measures, and data collection procedures.

Evaluation Capacity Building

The evaluation team and partners had trained the Women's Federation staff to administer interviews to IDUs, collecting participants' social and demographic information and outcome measures, including history of drug use and HIV risk factors. Research staff were trained to conduct an exit survey to check which VCT services were delivered and measure clients' satisfaction with the services. Counselors were trained to record information from the counseling sessions pertaining to HIV-testing results and the participant's initial reaction. The Women's Federation held weekly staff meetings to periodically review the implementation data and exchange counseling experience to improve the program.

Using Program Theory to Check for Potential Vulnerabilities

Two VCT sites were established in the community. They were renamed Health Education and Counseling Service to avoid any association with IDU

or HIV/AIDS. The name change arose from evaluators' discussion of the program theory with program staff. During one of the planning meetings, evaluators raised client safety issues to the Women's Federation staff. Because drug abuse is illegal in China, the evaluation team was concerned that IDUs might get arrested when they entered and left VCT sites. The program director and staff agreed with the evaluators' concern. To further ensure clients' safety, the director met with the mayor of the city to discuss this problem. The mayor agreed to tell local police chiefs not to interfere with the VCT sites' services and to abstain from monitoring or arresting IDUs during or after services. The Women's Federation also provided clients with a hotline number to call in case they were arrested because of receiving services. The strategy worked. The follow-up interviews indicated that no clients were arrested due to receiving VCT services.

Conclusive Assessment

The research design for the conclusive assessment consisted of the following components:

Inquiry Into the Process of Contextualizing the Intervention in the Community

The evaluation team interviewed the directors and counselors about adjuvants. They explained that to appropriately implement VCT with Chinese citizens, implementers would need to put VCT in the context of Chinese culture. Chinese culture stresses social relationships and collective values, while VCT mainly focuses on providing knowledge and skills related to HIV prevention. They decided to add two adjuvants to VCT and adjust the VCT protocol slightly, as follows.

Social Connection Adjuvant

The original protocol of VCT called for a short introduction session, focusing on confidentiality protection protocols, and then went straight into VCT counseling and testing procedures. Counselors were worried that this process would not work in China, where social relationships are essential to establishing effective communication. Chinese individuals generally are not receptive to advice and information provided by strangers. Clients would be very uncomfortable if counselors immediately engaged

(Continued)

(Continued)

in counseling conversations. The new component required counselors to first build a social connection with their clients. Counselors often used a client's neighborhood, schooling, and family origin as a starting point to build this relationship. In addition, counselors used their community roots and service record to enhance clients' trust.

Collectivity-Value Adjuvant

The VCT protocol stressed providing HIV-prevention education to clients during counseling sessions. Counselors believed the intervention would work better if they also emphasized the benefits of HIV-prevention education to clients' family members, the broader community, and the nation. This element was added to the counseling strategy.

Modifying a Demonstration Tool

Evaluators concluded that the VCT program had been implemented with high fidelity, with the exception of one modification. The original protocol called for the use of a realistic penis model to demonstrate and practice condom use. However, due to the conservative nature of Chinese society, the group was concerned that some IDUs would be embarrassed or offended by the use of the penis model. Counselors decided to use a banana as the model for clients to practice condom use on.

The evaluation team also conducted an environmental scan to identify potential events in the community that could affect outcomes. They looked at local news outlets, conducted key informant interviews, and held focus groups with participants. There was no need to consider events as part of the mesological intervention.

Using a Relatively Unobtrusive Quantitative Design to Address Biases and Assess Change

This study applied the one-group pretest-posttest design to assess the joint effects of the mesological intervention. IDUs living in a city in the province of Guangxi, in southern China, were targeted for VCT program recruitment. Potential participants were recruited by peer recruiters. A total of 226 IDUs agreed to participate in the study. Participants' HIV risk factors were discussed before they received VCT services. The participants were also

invited to be a part of a follow-up assessment of HIV risk factors 6 months after the intervention.

Using an Auxiliary Design to Triangulate Evidence

The auxiliary design consisted of the following elements:

Qualitative Evaluation

Evaluators also conducted two focus group meetings with clients. Participants in the meetings expressed satisfaction with the services provided and shared behavioral changes they had made to avoid contracting HIV. The qualitative findings were consistent with the evidence found in the quantitative design.

Pattern Matching

The VCT project used the nonequivalent dependent variable pretest-posttest design (Cook & Campbell, 1979; Shadish et al., 2002) to assess change in clients' HIV knowledge and risk behaviors. The basic structure of this design is similar to the one-group pretest-posttest design. However, in addition to the designated dependent variable, additional nonequivalent dependent variables are included for investigation purposes. The nonequivalent dependent variables must have the following two characteristics: Both designated and nonequivalent variables must be confronted with a similar set of threats to internal validity, and the intervention, theoretically, must affect only the designated dependent variables and not the nonequivalent variables. If the data support theoretical patterns, the evidence supporting the effect of the designated dependent variables is stronger than the evidence provided by the one-group pretest-posttest design, because threats to internal validity, such as maturation and reactivity to testing, have been mitigated (Cook & Campbell, 1979).

In the above study, the designated dependent variables were HIV knowledge, safe drug use, and protected sex. The nonequivalent dependent variable used in the study was cigarette smoking. A great majority of drug users in China are smokers. However, since VCT does not target cigarette smoking, theoretically, VCT should change HIV knowledge and HIV risk behaviors but not lead to a decrease in cigarette smoking. The finding provided additional evidence indicating that the program was effective.

(Continued)

(Continued)

Conclusions of the Evaluation

The evaluation team conducted a constructive assessment to ensure that the program had a coherent structure and the organization had sufficient capacity to participate in evaluation. A conclusive assessment was carried out after the constructive assessment. The constructive assessment identified several adjuvants used by stakeholders to make VCT work in their community. A comparison of the pretest and posttest behaviors in the quantitative design showed that the mesological intervention (VCT and adjuvants) increased participants' HIV knowledge and reduced risky behaviors. The auxiliary design provided additional evidence to support this claim. Data from pattern matching showed that the mesological intervention did not reduce participants' smoking behaviors, as expected. The qualitative data also indicated that participants had positive experiences with the intervention and their behavioral changes. In general, the findings indicated the program was effective.

CHECKLIST FOR RANKING REAL-WORLD EVALUATIONS

The above discussion clearly indicates that the experimentation evaluation approach and holistic effectuality evaluation approach are very different in terms of priorities, principles, and methodologies. The experimentation evaluation approach rates research designs in a hierarchy of quality, with experimental designs being best, quasi-experimental designs coming next, and pre-experimental designs being the least preferred. The holistic effectuality evaluation approach judges the quality of an evaluation in terms of two components: constructive assessment and conclusive assessment. This textbook proposes the checklist and scale from 0 to 10 shown in Table 11.1 to rate the quality of a real-world evaluation.

In the rating system presented in Table 11.1, a real-world evaluation receives 0.5 points for each of the following components of the constructive assessment: conducting a constructive outcome evaluation to ensure the coherence of the program, assessing and building evaluation capacity as needed, conducting a participatory evaluation, and using program theory or logic models to work with stakeholders to check whether the program has vulnerabilities.

Table 11.1 Rating System for the Holistic Effectuality Evaluation Approach*Constructive assessment* 0–2

Conducting a constructive outcome evaluation to ensure the coherence of the program	0–0.5
Assessing and building capacity as needed	0–0.5
Conducting a participatory evaluation	0–0.5
Using program theory or logic models to work with stakeholders to check whether the program has vulnerabilities	0–0.5

Conclusive assessment 0–8

Inquiring into the process of contextualizing the intervention in the community	0–2
Using an unobtrusive quantitative design to address biases and assess change	0–3
Using an auxiliary design to triangulate evidence	0–1
Replicating the mesological intervention*	0–2

NOTE: Meaning of point totals—Outstanding: 9–10; Excellent: 7–8; Satisfactory: 5–6; Unsatisfactory: 0–4

*This activity is optional.

In the conclusive assessment, the points for inquiring into the process of contextualizing the intervention in the community range from 0 to 2, depending how thorough the inquiry is. The score earned by using a relatively unobtrusive quantitative design to address biases and assess change ranges from 0 to 3: An evaluation that uses a quasi-experiment will get a score of 3, the one-group pretest-posttest design earns a 2, and the cross-section comparison gets a 1. The auxiliary design is scored either a 0 or 1, depending on whether an auxiliary design is used. The rating system adds 1 point for replicating the mesological evaluation at another site and 2 points for replicating at multiple sites. This item is optional, according to the holistic effectuality evaluation. A total score of 9–10 is considered outstanding; 7–8, excellent; 5–6, satisfactory; and 0–4, unsatisfactory.

The above ranking system indicates that a given evaluation could be rated as low quality according to the experimentation evaluation approach but high quality according to the holistic effectuality approach. For example, the experimentation

evaluation approach would consider the VCT example above a low-quality evaluation because of the use of the one-group pretest-posttest design. However, the holistic effectuality evaluation perspective would give the evaluation a score as high as 8, rating it as an excellent real-world evaluation.

USEFULNESS OF THE HOLISTIC EFFECTUALITY EVALUATION APPROACH

The holistic effectuality evaluation approach is useful for evaluators and stakeholders in the following areas:

Providing Theory and Methodology for Real-World Outcome Evaluation

Traditionally, only the experimentation evaluation approach has been available to guide outcome evaluation. When evaluators and stakeholders are interested in evaluating an intervention program's real-world effects, they have had no choice except to follow the experimentation evaluation approach. As a consequence, adjuvants, in which stakeholders are greatly interested, are either ruthlessly eliminated or simply neglected in the evaluation. The holistic effectuality evaluation approach addresses this problem by providing an underlying theory and a methodology for conducting real-world outcome evaluation. Evaluators and stakeholders now have two options. If they are interested in evaluating an intervention's pure independent effects, they can select the experimentation evaluation approach. If they are interested in evaluating an intervention's real-world effects, they can select the holistic effectuality evaluation approach.

Providing Insight Into the Relationship Between Adjuvants and Internal Validity

The holistic effectuality evaluation approach provides new insights into threats to internal validity. It argues that if evaluators want to assess an intervention's pure independent effects, then they should control for them, as with the experimentation evaluation approach. However, if evaluators and stakeholders want to assess an intervention's real-world effects, the concept of threats to internal validity is likely misleading. If evaluators blindly rule to

eliminate such factors from the analysis, the resulting evaluation becomes irrelevant to real-world practice. The holistic effectuality evaluation approach, by using the concept of adjuvants, allows evaluators to understand the additional efforts made by practitioners and clients to make an intervention work and evaluate the mesological program accordingly. By examining the relationship between an intervention and its adjuvants, evaluators are more likely to generate results that are relevant to stakeholders' practice. Furthermore, this kind of information may also be useful for members of the academic community when they are developing ideas for innovative intervention programs to be applied to real-world settings.

Inspiring Evaluators to Develop Indigenous Evaluation Theories and Methodologies

The experience of constructing the holistic effectuality evaluation approach illustrates how program evaluation can be enriched by developing innovative evaluation theories and methodologies that build on and extend preceding work. As discussed in this chapter, imported theories and methodologies are useful for program evaluation up to a certain point. Beyond that, evaluators must develop their own theories and methodologies to advance the field.

THE EXPERIMENTATION EVALUATION APPROACH VERSUS THE HOLISTIC EFFECTUALITY EVALUATION APPROACH

Chapters 10 and 11 indicate that there are two ways to conduct outcome evaluation: the experimentation evaluation approach and the holistic effectuality evaluation approach. Since each approach has its own purpose, focus, principles, priorities, and methodologies, it would be inappropriate to claim that one approach is universally better than the other. Evaluators and stakeholders first must understand their evaluation needs and only then select an appropriate methodology.

As a rule of thumb, if evaluators and stakeholders are interested in assessing an intervention's pure independent effects, they should choose the experimentation evaluation approach. If they are interested in an intervention's real-world joint effects, they should select the holistic effectuality evaluation approach. As a guide to selecting the best approach, the strengths and limitations of each type of effect are summarized here.

Pure Independent Effects

Strengths

Information about the pure independent effects of an intervention shows precisely how much an intervention, by itself, affects an outcome. This kind of effect is consistent with the common standard for research literature because it explores whether a causal relationship exists between an independent variable and a dependent variable. One of the benefits of this kind of evaluation is that prestigious designs such as RCTs or other experimental methods can be used. From the traditional research perspective, these experimental methods provide rigorous evidence. Researchers view this kind of information as essential for developing scientific knowledge and thus favor an assessment of an intervention's pure independent effect.

Limitations

An intervention's pure independent effect may not be relevant to real-world practice. For example, the application of RCTs to assess the Zumba program would obtain the intervention's pure independent effect on participants' weight loss. Unfortunately, when an RCT blindly rules out all threats of internal validity, it also eliminates the potential contributions of adjuvants to the program that are essential for real-world effects. Thus, there is a risk that a pure independent effect will be irrelevant to the effect an intervention will achieve in the real world.

Another limitation of RCTs or other experimentation methods is that they are highly time-consuming and resource-intensive. These days, technology and community needs change rapidly; the lengthy time frame required to conduct RCTs may not meet stakeholder demands. The high price tag of RCTs is another barrier for community-based organizations that wish to assess a pure independent effect.

Real-World Joint Effects

Strengths

Evaluating an intervention's real-world effect allows evaluators to understand what actually happens in a community and assess the benefits clients gain from the intervention program. Stakeholders are usually interested in knowing what clients are actually getting. For example, a real-world evaluation of my

participation in the Zumba class would find out I lost 20 pounds. However, as discussed previously, this weight loss was a joint effect, resulting from the combination of the Zumba class, the instructor's interest in the participants' well-being, my strong motivation to lose weight, the incorporation of high-impact steps into my choreography, an increase in the number of classes I attended, and changes to my diet. The real-world effect was very important to me and my instructor.

Furthermore, stakeholders' are likely to view the evaluation as highly relevant to their work. Moreover, real-world effects may be more easily transferred to other communities. An evaluation that assesses real-world effects is also relatively less time-consuming and resource-intensive than is an evaluation of pure independent effects.

Limitations

An evaluation of an intervention's real-world effects usually cannot answer the question of exactly what contribution was made by the intervention alone. For example, as discussed previously, I can definitely say the Zumba class was the driving force behind my 20-pound weight loss. However, from a researcher's viewpoint, this statement is not precise; researchers prefer an evaluation that can estimate how many pounds of weight loss occurred due to the Zumba class alone.

BUILDING EVIDENCE FROM THE GROUND UP

Since pure independent effects and real-world joint effects each make contributions to understanding interventions, those seeking to solve a community problem ideally should have access to knowledge about both types of effects. Unfortunately, credible evidence is defined by the evidence-based practice movement as only evidence produced by RCTs (Nutbeam, 1999; Speller, Learmonth, & Harrison, 1997; Stephenson & Imrie, 1998; Tilford, 2000). This tradition mainly focuses on evidence that rules out all threats to internal validity. In other words, only evaluations of pure independent effects are regarded as credible. As a consequence, the scope of current evidence-based practice excludes evidence of the real-world joint effects that are greatly interesting to practitioners and other stakeholders. This book uses the term *evidence gap* to describe the absence of information available to stakeholders resulting from the usual evaluation of an intervention's pure independent effects and the grossly neglected evaluation of real-world effects.

Since stakeholders are much more interested in interventions' real-world effects than the pure independent effects found in research settings, the evidence gap presents a huge obstacle to communication between the academic and practice communities as they seek consensus on evidence-based practices and ways to collaborate. The holistic effectuality approach proposes a way to build evidence from the ground up. Real-world programs with evidence as to their effectiveness are more generalizable than interventions deemed efficacious. Furthermore, intervention programs determined to have real-world effectiveness could be further examined by RCTs to sort out which component or components of the program have pure independent effects on outcomes. Building evidence from the ground up may not only make possible the kind of evidence-based practices of interest to stakeholders but also enhance communication and collaboration between researchers and stakeholders, allowing them to more effectively solve community problems. These issues will be discussed in more detail in Chapter 15.

QUESTIONS FOR REFLECTION

1. Describe evaluators' dilemma regarding the application of the experimentation evaluation approach to outcome evaluation. Discuss how evaluators currently cope with this dilemma.
2. Describe efficacy evaluation and effectiveness evaluation. Why does the academic community favor efficacy evaluation? If you are a practitioner, what is your view on this issue?
3. Discuss why it is important to differentiate the Campbellian validity typology from the experimentation evaluation approach.
4. Discuss why the holistic effectuality evaluation approach argues that real-world outcome evaluation should be a hybrid (constructive and conclusive) evaluation rather than only a conclusive evaluation, as assumed by the experimentation evaluation approach.
5. Discuss the relationship between stakeholders and evaluators in the context of both a conclusive outcome evaluation and a hybrid evaluation.
6. Discuss why commonly perceived threats to internal validity were not necessarily threats in the Zumba example. In what ways were they actually beneficial to understanding the program?

7. Explain the concept of adjuvants and describe the types of adjuvants and how they relate to the various threats to internal validity. Give examples of each type of adjuvant.
8. Discuss why adjuvants are important to stakeholders' practice. Why might it be undesirable to treat adjuvants as threats to internal validity and eliminate them from consideration in an evaluation.
9. Define the terms *biases* and *adjuvants*.
10. What is a joint effect? Discuss why real-world programs tend to have joint effects rather than pure independent effects.
11. Discuss the components of constructive assessment in a real-world evaluation. Why are they important?
12. Discuss the components of conclusive assessment in a real-world outcome evaluation? Why are they important?
13. The holistic effectuality evaluation approach argues that the one-group pretest-post-test design is not a weak design for a real-world evaluation. This view runs counter to the principles of the experimentation evaluation approach. Discuss your position on this issue and explain your position.
14. Select an article from the literature that describes a real-world outcome evaluation and apply the rating system provided by the holistic effectuality evaluation approach to rate the quality of the article. Discuss your findings. Do you agree with the rating? Why or why not?
15. Discuss the strengths and limitations of assessing an intervention's pure independent effects versus assessing its joint effects. Which might you assess in your next evaluation. Why?

CHAPTER 12



The Theory-Driven Approach to Outcome Evaluation

The major question asked in either the experimentation evaluation approach or the holistic effectuality evaluation approach is whether an intervention program works or not. However, if stakeholders and evaluators desire to understand both the merits of a program *and* how its transformation processes can be exploited to improve the intervention, then evaluators need to apply the theory-driven approach to upgrade a typical outcome evaluation into a theory-driven outcome evaluation. Theory-driven outcome evaluation takes into account both underlying causal mechanisms and the implementation process when assessing the effect of a program; it is a hybrid evaluation and can provide stakeholders with an understanding of whether a program is reaching its goals and document insightfully the hows and whys of program success or failure (Bickman, 1990; Chen, 1990, 2005, 2012a, 2012b; Coryn et al., 2011; Weiss, 1997). This is why theory-driven outcome evaluation is placed in the evaluation typology with the enlightenment assessment strategy (see Table 2.1): It enlightens stakeholders as to the crucial assumptions under which their program is expected to operate in the field each day and the program components that contribute to, and hinder, overall success. This chapter will show how a typical outcome evaluation as discussed in Chapter 10 or 11 can, using program theory, be upgraded to theory-driven outcome evaluation. Neither efficacy nor effectiveness evaluation has to follow the black-box evaluation route.

The advantages cited here for theory-driven outcome evaluation—as well as other advantages—are well documented in the literature (e.g., Bickman, 1990; Chen, 1990, 2005, 2012a, 2012b; Donaldson, 2003, 2007; Nkwake, 2013; Weiss, 1997). Nevertheless, a brief review of the three most significant advantages may be helpful.

1. *Theory-driven outcome evaluation serves both accountability and program improvement needs.* In general, so-called black-box evaluation fulfills accountability requirements only. The information provided by theory-driven outcome evaluation can be used to serve accountability and program improvement needs alike because this type of evaluation generates *two* kinds of information for stakeholders. Initially, it assesses whether a program is achieving its predetermined goals, thereby meeting stakeholders' need for accountability; in this, it is like black-box evaluation. Subsequently, it investigates why and how a program succeeds or does not succeed, helping stakeholders in the task of better understanding and improving their programs. By exploring underlying causal mechanisms, theory-driven outcome evaluation has the potential to contribute substantively to the program to which it is applied and to the field of evaluation in general.

2. *Theory-driven outcome evaluation can comment on construct validity.* Construct validity measures the degree to which an outcome evaluation assesses the right thing in the right way. As explained in Chapter 1, it is sometimes possible for programs to achieve their goals via several different channels. Some of these channels, however, may be illegitimate or may lack social approval. A program that achieves its effect in illegitimate or socially disapproved ways can hardly be termed a success. A black-box evaluation performed on a program that succeeds via unsanctioned channels will ignore underlying causal mechanisms and so miss this obvious demerit. Thus, the black-box evaluation may lack construct validity. Theory-driven outcome evaluation has higher construct validity because by examining the causal mechanisms underlying a program, it provides evidence that the evaluation is assessing the right thing. For example, say an education program claims that it can increase students' reading scores by enhancing students' motivation to read. We will be more confident in an outcome evaluation that shows that, indeed, students' motivation improved and students' reading scores increased than in an evaluation that shows only that reading scores increased while providing no information on whether students' motivation improved.

3. *Theory-driven outcome evaluation can increase internal validity.* A theory-driven outcome evaluation requires the specification of program theory. The process of specifying program theory gives the evaluation the advantage of pattern matching, as discussed in Chapter 11. As pointed out by Trochim (1998), Pawson and Tilly (1997), and Chen (1990), elaboration of theoretical patterns can enhance internal validity.

CLARIFYING STAKEHOLDERS' IMPLICIT THEORY

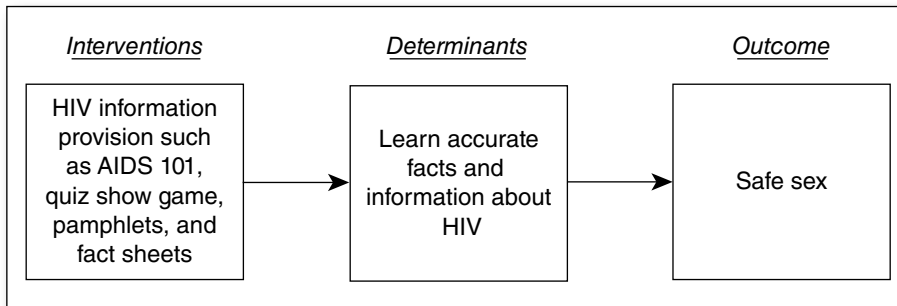
Making Stakeholder Theory Explicit

Chapter 3 explained that when formal theory is the basis of a program, that theory usually specifies appropriate determinants. For example, if a program is based on the health belief model (see Kohler, Grimley, & Reynolds, 1999, as an illustration), it is “assigned” the determinants *perceived susceptibility*, *perceived severity*, *perceived benefits*, *perceived barriers*, and *efficacy expectations*; each must be included in any model of the program’s intervening mechanisms.

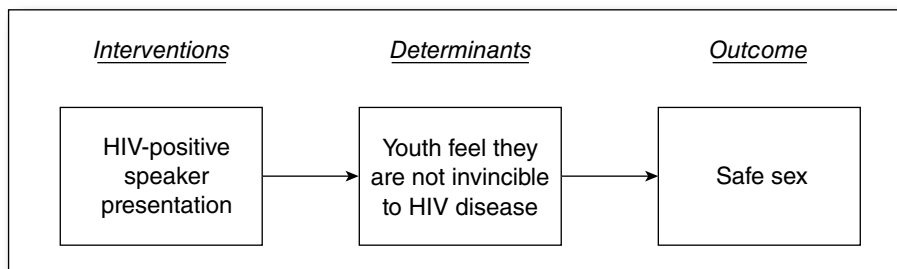
Because a formal theory is, in such cases, already in place, designing and conducting theory-driven outcome evaluation of the program is a straightforward prospect. Admittedly, however, most programs do not spring from well-developed formal theory. The majority have their sources in stakeholder theories, a fact that causes critics to believe that evaluators will meet with insurmountable obstacles should they attempt to help stakeholders clarify (or develop) their theories. However, the Laub, Somera, Gowen, and Diaz (1999) study offers a concrete, highly persuasive example that shows evaluators can overcome these purported obstacles. Laub’s team considered HIV-prevention programs for youth conducted by community-based organizations (CBOs). The great majority of these are based upon stakeholder theory, not scientific theory. Interventions popular with CBOs for HIV prevention are classes, HIV-positive speakers, engaging youth in volunteer work for people with HIV/AIDS, condom distribution, and outreach. Laub and colleagues traced the connections between each of these popular options and implicit underlying stakeholder theory and made this theory explicit.

For example, one set of interventions documented by the Laub et al. (1999) study provided HIV-prevention information in the form of an AIDS 101 class, a quiz show game, pamphlets, and fact sheets. These interventions did not come from nowhere but rather arose from stakeholders’ theory. The implicit theory at work through these interventions is that youth have unsafe sex because they do not have accurate information about HIV. Accordingly, if young people are provided with accurate information, they are more likely to practice safe sex. Stakeholders’ implicit theory can be made explicit, as illustrated in Figure 12.1a.

Another example is the presentation by an HIV-positive speaker. Stakeholders’ implicit theory supporting such an intervention is that because youth have

Figure 12.1a Implicit Theory for HIV-Prevention Information Intervention

SOURCE: Adapted from Laub, Somera, Gowan, & Diaz (1999).

Figure 12.1b Implicit Theory for Presentations by HIV-Positive Speakers

SOURCE: Adapted from Laub, Somera, Gowan, & Diaz (1999).

never known a person their age with HIV, they feel invincible. Again, stakeholders' implicit theory can be made explicit, as illustrated in Figure 12.1b: Stakeholders' theory is that presentations by HIV-positive speakers will demonstrate to youth that they are not immune to HIV disease, and as a result these young people will practice safer sex.

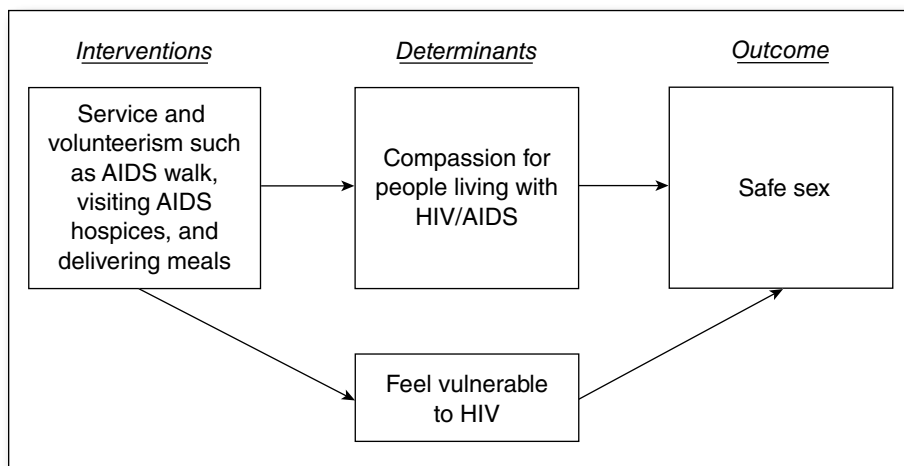
A third example is service and volunteerism. These interventions included participating in an AIDS walk, visiting AIDS hospices, and delivering meals to people with AIDS. Stakeholders' implicit theory for this intervention is

that youth lack compassion for people living with HIV/AIDS and do not feel vulnerable to HIV. This implicit theory can be made explicit, as illustrated in Figure 12.1c: Stakeholders believe that service and volunteerism can increase youths' compassion for people living with HIV/AIDS and show them that they are vulnerable to HIV, which, in turn, will increase their practice of safe sex.

Laub et al. (1999) argued that making a stakeholder theory explicit has another advantage for stakeholders: They become aware of the potential limitations of an intervention. For example, when an intervention provides information about HIV, the theorizing process might indicate that accurate information alone might not be sufficient to motivate safe-sex behavior. This kind of insight is useful for stakeholders to consider as they add components to strengthen their program.

As suggested in Chapters 3, 4, and 5, evaluators can use interviews or working group meetings to help stakeholders develop their theories specifically and explicitly and, especially, to identify determinants in their change models. For a small program, an evaluator and stakeholders who have prepared well may need only one or two meetings to accomplish this. For a large program, more meetings will probably be needed. When implicit theories have been clarified or

Figure 12.1c Implicit Theory for Service and Volunteerism Intervention



SOURCE: Adapted from Laub, Somera, Gowan, & Diaz (1999).

made explicit, specifying the change model used for theory-driven outcome evaluation is easy.

Building Consensus Among Stakeholders Regarding Program Theory

Agreement among stakeholders about program theory is often not difficult to reach. However, even if some components of the program theory do spark disagreement among key stakeholders, this is not necessarily an obstacle to evaluation. Rather, disagreement may indicate a need for the evaluation to test stakeholders' competing hypotheses. For example, if stakeholders disagree on which determinant mediates the program's intervention and its outcome, the evaluator could gather empirical data on the relative importance of competing determinants and provide that feedback to stakeholders. In the example depicted in Figure 12.1c, key stakeholders might disagree on what mediates the program's intervention (volunteer activities) and the program's outcome (safe sex). Two competing determinants are considered: increased compassion for people living with HIV/AIDS and the participants' elevated feelings of susceptibility to HIV. Evaluators could include both hypotheses in the evaluation and provide empirical data that would demonstrate the relative importance of both.

GUIDELINES FOR CONDUCTING THEORY-DRIVEN OUTCOME EVALUATION

Three general guidelines are available to evaluators who are designing and conducting a theory-driven outcome evaluation.

1. *Establish a common understanding between stakeholders and evaluators of what theory-driven outcome evaluation is and what it does.* Both parties must agree on the need for theory-driven evaluation, the steps to be used in the evaluation, and the defined roles each party will play in the evaluation. The stakeholder-evaluator dialogue should specifically answer the following questions to everyone's satisfaction: What will the evaluative information be used for? Do stakeholders want an efficacy evaluation or an effectiveness evaluation? What is the program theory? What procedures could be used to clarify the program theory or help stakeholders develop one?

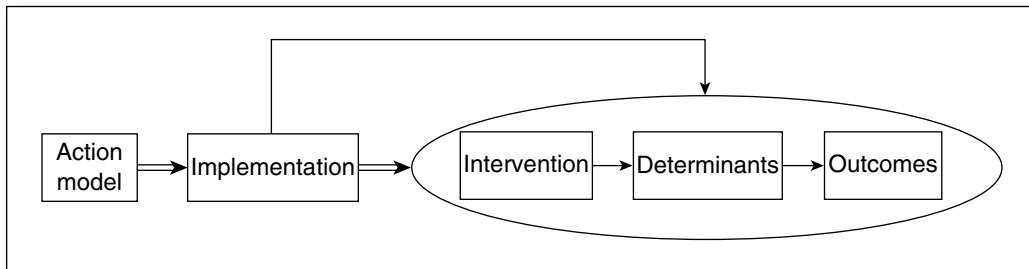
2. *Clarify the stakeholders' theory.* Program theory is the foundation of a theory-driven outcome evaluation. Evaluators could use the conceptualization

facilitation approach discussed in Chapters 3, 4, and 5 to clarify stakeholders' program theory or help them develop one. An early question for stakeholders and evaluators is whether evaluation should focus on a change model, an action model, or both. The answer depends on the extent of evaluative information desired by the stakeholders. Typically, theory-driven outcome evaluation focuses on the change model (program rationale) by examining the relationships among intervention, determinants, and outcomes. At times, however, program stakeholders wonder how program components (i.e., type of implementers, delivery models, clients) relate to the intervention-outcome relationship. Such an inquiry relies on findings about an action model (program plan), a topic discussed in Chapter 5. Alternatively, both the change model *and* the action model come to the forefront when stakeholders express a need for outcome evaluation combined with process evaluation. These different types of theory-driven outcome evaluation are detailed in the remainder of this book.

3. *Construct a research design.* Theory-driven outcome evaluation demands ample contextual information and therefore tends to rely on mixed methods of data collection. The framework of program theory meaningfully links quantitative and qualitative data in order to generate a comprehensive view of a program (Chen, 1996). Due to its comprehensiveness, a theory-driven outcome evaluation typically relies on the collection of data beyond that common for a black-box evaluation. For example, in preparation for assessing a change model, an evaluator conducting a theory-driven outcome evaluation would need to gather data on the intervention and outcome as well as on the determinant. The need to collect additional data can mean increased costs, but these are minimized when theory-driven outcome evaluators—fulfilling their usual tasks of site visits, client surveys, or other field-based collection of intervention and outcome data—arrange to collect the additional information about determinants concurrently.

TYPES OF THEORY-DRIVEN OUTCOME EVALUATION

Based upon the conceptual framework of program theory, as shown in Figures 3.1 and 3.2, the basic elements of theory-driven outcome evaluation are illustrated in Figure 12.2. The illustration makes clear that the core of theory-driven outcome evaluation is assessment of relationships among the intervention, the determinants, and the outcomes. These relationships, however, are influenced by the implementation of the action model, that is, by whether that implementation is of poor or high quality. For example, when implementers are incompletely committed to the intervention, it will probably not much affect the determinants

Figure 12.2 Elements of Theory-Driven Outcome Evaluation

or outcomes. Figure 12.2 represents the two links between program implementation and the change model. One, indicated by the single-line arrow, is the activation of the intervention by the implementation; the implementation delivers the intervention to clients. The other link, indicated by the double-line arrow, is the implementation process's shaping of causal relationships among variables in the change model.

To be more specific, two kinds of causal mechanisms may underlie a program: mediating and moderating. A *mediating causal mechanism* is a component of a program that intervenes in the relationship between two other components. Figure 12.2 illustrates a determinant mediating the relationship between an intervention and an outcome. In a relationship of this type, the intervention cannot affect the outcome unless it also affects the determinant. But simply affecting the determinant does not guarantee the intervention's success, which also requires the determinant's ability to change the outcome. Thus, selection of the appropriate determinant is central to a program's performance. For example, consider a program offering HIV-prevention classes to female migrant workers. This intervention suggests that program designers believe that female migrant workers court risk of infection because they lack knowledge of HIV/AIDS and the skills to avoid the virus. Education—the class—is the program's intervention. Its determinants are HIV/AIDS knowledge and prevention skills. Its measurable outcome is condom use. Thus, to be effective, the educational intervention must alter the women's knowledge and skills; furthermore, the program designers' belief that lack of knowledge and skills causes the women's high-risk behavior must be valid. Again, the determinant mediates, or is an intervening variable in, this causal process.

The second type of causal mechanism—the *moderating causal mechanism*—represents a relationship between program components that is enabled, or conditioned, by a third factor. In the presence of this third factor, the two

components' relationship is manifested. In its absence, this relationship dissolves. Returning to the HIV-prevention program to illustrate, assume that stakeholders suspect that their program's effect on the female migrant workers hinges on the cultural backgrounds of the women's sexual partners. The stakeholders worry that if those backgrounds prize male dominance, the intervention will accomplish nothing. On the other hand, if the culture does not embrace male dominance, the intervention will likely succeed.

When disparate parts of the framework of the causal mechanism are isolated (see Figure 12.2), at least three kinds of theory-driven outcome evaluations can be designed: (a) an intervening mechanism evaluation approach, which focuses on the mediating process; (b) a moderating mechanism evaluation approach, which focuses on the moderation process; and (c) an integrative process/outcome evaluation approach, which focuses on the linkage of the intervention to outcomes via implementation and causal processes.

THE INTERVENING MECHANISM EVALUATION APPROACH

The intervening mechanism evaluation approach assesses whether the causal assumptions underlying a program are functioning as stakeholders had projected (Chen, 1990). To date, intervening mechanism evaluation is the most popular application of theory-driven outcome evaluation (Donaldson, 2003; Mark, 2003). It is not always labeled in the same way by those who apply it. Some evaluators have referred to it as "theory of change evaluation" (Connell, Kubisch, Schorr, & Weiss, 1995) or "theory-based evaluation" (Rogers, Hasci, Petrosino, & Huebner, 2000; Weiss, 1997).

The basic task of the intervening mechanism evaluation is to assess the change model portion of the program theory. The intervening mechanism is modeled in Figure 12.3.

Figure 12.3 The Basic Model of Intervening Mechanism Evaluation

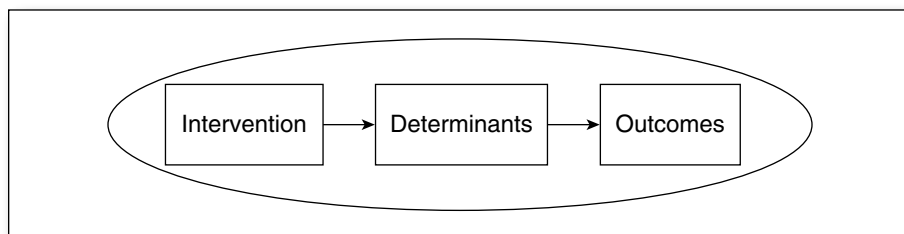


Figure 12.3 indicates that the major difference between intervening mechanism evaluation and traditional efficacy and effectiveness evaluation is the former's inclusion of the determinants that intervene between the intervention and outcome. An example of how the intervening mechanism evaluation has been used in research is presented to acquaint readers with this approach and simplify the remaining discussion. Donaldson, Graham, and Hansen (1994) evaluated the Adolescent Alcohol Prevention Trial. This project focused on the three determinants said in the literature to be important causes of adolescents' drug use: skills for resisting drug use, perception of the prevalence of drug use, and cognitive acceptance of drug use. Resistance skills (i.e., the behavior involved in turning down drug offers) were taught to adolescents enrolled in the program. In addition, a normative education program helped correct the participants' erroneous perceptions about the prevalence and acceptance of adolescent substance use and sought to establish conservative norms. The evaluation design incorporated testing of the two interventions' ability to affect the determinants. It also assessed whether the determinants had the power to affect three outcome variables: alcohol use, cigarette use, and marijuana use. Figure 12.4 shows the change model for the Adolescent Alcohol Prevention Trial.

Data from the intervening mechanism evaluation showed that normative education was employed successfully in the program; it activated beliefs concerning the prevalence and acceptance of drug use, which in turn reduced drug use. The data also showed that, although resistance training did strengthen the adolescents' resistance skills, possessing such skills did not affect drug use. The authors argued that their results strongly support the theoretical underpinnings of normative education interventions. Their study demonstrates the kind of information that theory-driven outcome evaluation provides: It answers the question of whether an intervention affects outcomes, and then it answers the question of *why* the intervention does or does not do so.

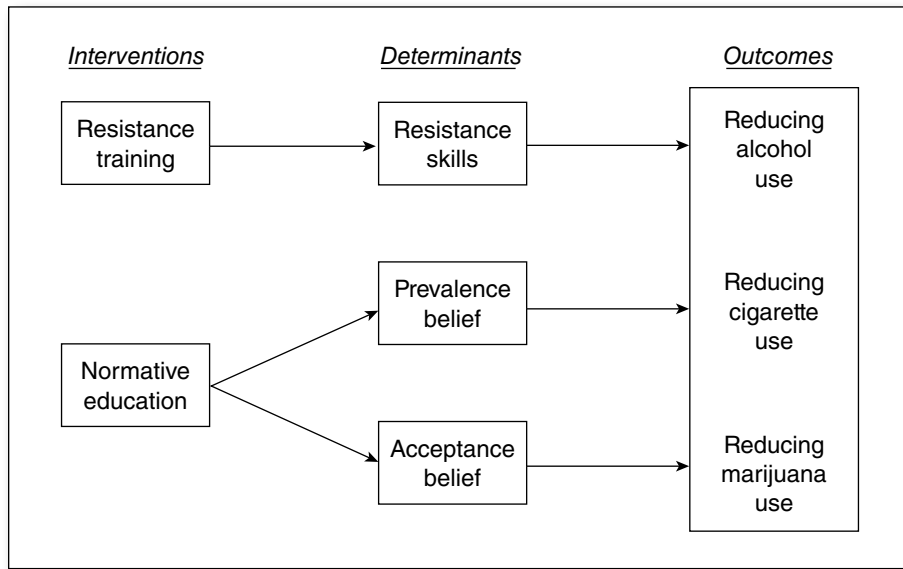
Two Models of the Intervening Mechanism Evaluation Approach

Two basic models of intervening mechanism evaluation predominate in the discipline: linear and dynamic.

The Linear Model

The linear model is currently a very popular application of intervening mechanism evaluation. Linear models assume that the causal relationships

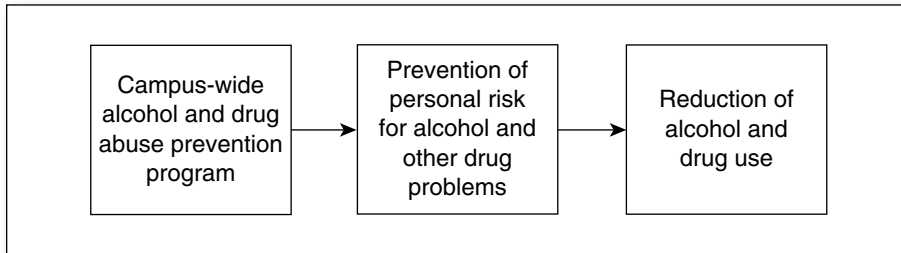
Figure 12.4 The Change Model for the Adolescent Alcohol Prevention Trial



SOURCE: Adapted from Donaldson, Graham, & Hansen (1994).

among interventions, determinants, and outcomes are unidirectional: intervention affects determinant, and determinant then affects outcome. No reciprocal relationships operate among the variables. In linear models, the number and sequence of the determinants under study determine the model's form. The causal diagrams in Figures 12.5, 12.6, and 12.7 illustrate the common linear model forms.

One-Determinant Model. This model, represented by Figure 12.5, contains a single determinant and is the fundamental model for intervening mechanism evaluation. The one-determinant model is illustrated here by an evaluation of an alcohol and drug abuse prevention program at a college (Miller, Toscova, Miller, & Sanchez, 2000). The intervention consisted of multiple components: print media, videotapes, speakers, referral services, and development of self-control. The determinant was perception of risk, and the outcome was a reduction in alcohol and drug use among the students on the campus where the program was established. As predicted, the data showed that after the interventions, there was

Figure 12.5 An Example of a One-Determinant Model

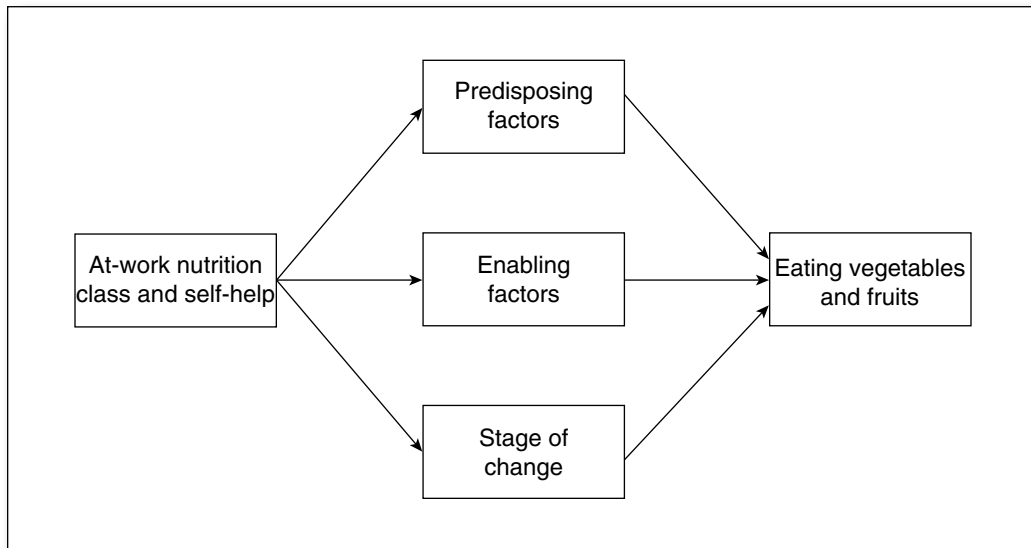
SOURCE: Adapted from Miller, Toscova, Miller, & Sanchez (2000).

heightened awareness on campus of the risks of substance abuse, which in turn reduced alcohol and drug use there. The one-determinant model is relatively easy to construct. For example, all of the stakeholders' implicit theories in HIV-prevention activities illustrated in Figures 12.1a, 12.1b, and 12.1c are one-determinant models.

Multiple-Determinant Model, No Sequential Order. Another common linear model includes two or more determinants, each affected by the intervention or affecting the outcome but in no particular sequence. A workplace nutrition program provides an example of the multiple-determinant model (Kristal, Glanz, Tilley, & Li, 2000). The intervention featured at-work nutrition classes and self-help. The stakeholders and evaluators selected three determinants: predisposing factors (skills, knowledge, belief in diet-disease relationship); enabling factors (social support, perceived norms, availability of healthful foods); and stage of change (action and maintenance stages being subsequent to the intervention). The outcome variable was dietary change (eating vegetables and fruits). The model of this program is illustrated in Figure 12.6.

Kristal and colleagues found that the intervention did enhance predisposing factors as well as the likelihood of entering and remaining in the subsequent stages of change. They also found, however, that the intervention did not affect enabling factors. The program was failing because the intervention was failing to activate one of the three determinants.

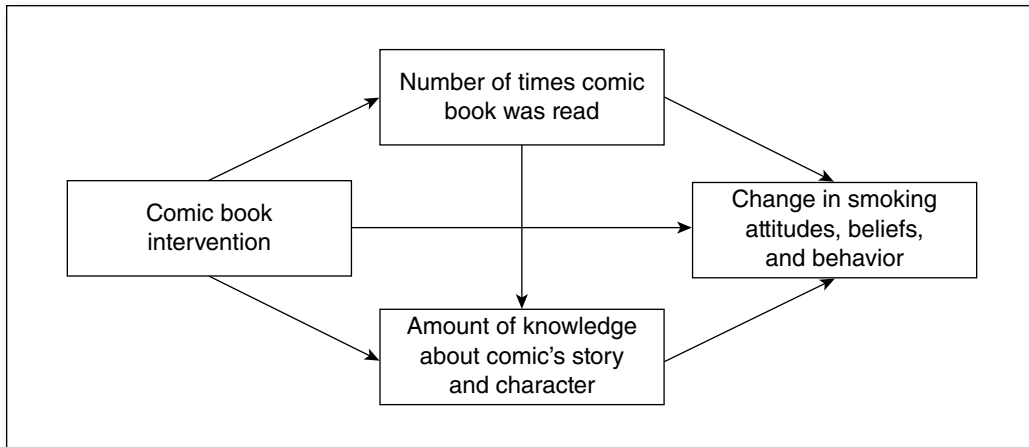
Multiple-Determinant Model, Sequential Order. This model contains two or more determinants aligned in a causal order. That is, certain determinants

Figure 12.6 Workplace Nutrition Program as a Multiple-Determinant, No Sequential Order Model

SOURCE: Adapted from Kristal, Glanz, Tilley, & Li (2000).

affect others in a particular sequential order. An example of this kind of linear model is found in an evaluation of a school-based antismoking campaign (Chen et al., 1988). The intervention contained components such as an anti-smoking comic book, discussions of the health messages the comic book delivered, and parental notification about the intervention program. The determinants of the model, in sequence, were the number of times the comic book was read and knowledge of the comic book's story and characters. The sequential order indicates that repeated reading of the comic book changed the extent of knowledge about the plot and characters. The sequence is illustrated in Figure 12.7.

The outcome to be measured was a change in attitudes, beliefs, and behaviors related to smoking. The evaluation determined that the distribution of the comic book affected the number of times the comic book was read, which in turn affected knowledge of its content. However, neither of these determinants was shown to affect students' smoking-related attitudes, beliefs, or behaviors.

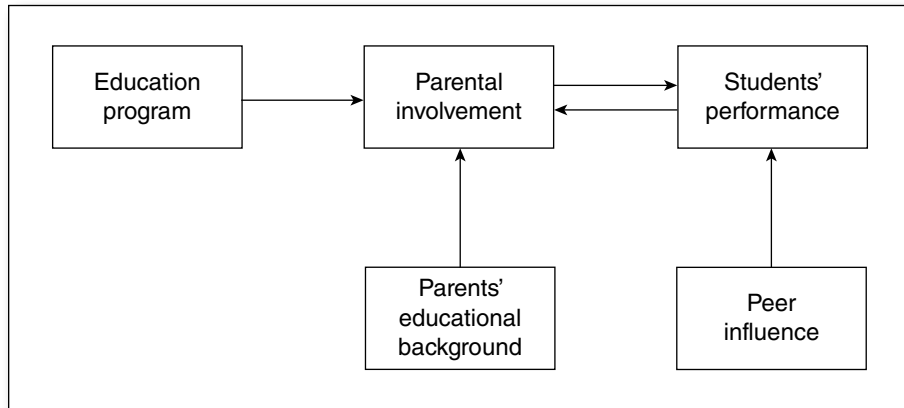
Figure 12.7 Antismoking Program as a Multiple-Determinant, Sequential Order Model

SOURCE: Adapted from Chen, Quane, & Garland (1988).

The Dynamic Model

The dynamic model of intervening mechanism evaluation assumes that multidirectional, reciprocal causal relationships exist among intervention, determinant, and outcome. The relationship between determinant and outcome, especially, is reciprocal rather than one-way: The determinant affects the outcome, and the outcome also affects the determinant. A hypothetical educational program illustrates the model well. The project's focus is to equip parents with skills and strategies to assist their children with homework; homework has been chosen as a determinant of primary students' school performance. The model makes clear, however, that the relationship between parental involvement and student performance need not be linear. Parents becoming more involved in a child's schoolwork might improve the child's performance; then, seeing the improved performance, parents perhaps might feel gratified and be stimulated to devote time and effort to remaining involved in the child's education. This form of the dynamic model is represented in Figure 12.8.

The dynamic model appears to be a sensible approach for many evaluation situations, but it is not widely applied at present by program evaluators. (The

Figure 12.8 Education Program as a Dynamic Model

literature does, however, widely discuss one case of intervening mechanism evaluation using the dynamic model—an evaluation of the Transitional Aid Research Project by Berk, Lenihan, & Rossi, 1980. This study has several implications for theory-driven evaluation, including those noted in Chen, 1990, and Shadish et al., 2002.) The reasons behind the limited use of the dynamic model to date include the high difficulty of constructing a model that can be assessed with the data available. The data analysis required in using a dynamic model in intervening mechanism evaluation relies on advanced statistical models such as the simultaneous equations model. These technical challenges aside, the dynamic model is a promising topic for future studies.

Some Theoretical Bases of the Intervening Mechanism Evaluation

Conducting an intervening mechanism evaluation usually requires both qualitative and quantitative research methods. In general, application of the intervening mechanism evaluation approach happens in two stages. In the model formulation stage, qualitative methods are essential for making explicit the stakeholders' change model. Methods such as the interview and the focus group are well suited to this task. In the data collection and analysis stage, quantitative methods become useful. Quantitative methods have a long tradition of

being used to collect the data needed to test models and analyze that data. For example, path analysis and structural equation modeling are well-established statistical techniques for testing models. The established nature of quantitative research methods is perhaps why most evaluators are given to applying these kinds of tests during an intervening mechanism evaluation. Note that this does not mean that qualitative methods cannot be useful to the evaluator here, but the question of the best ways to apply qualitative methods to testing change models is a topic for further investigation.

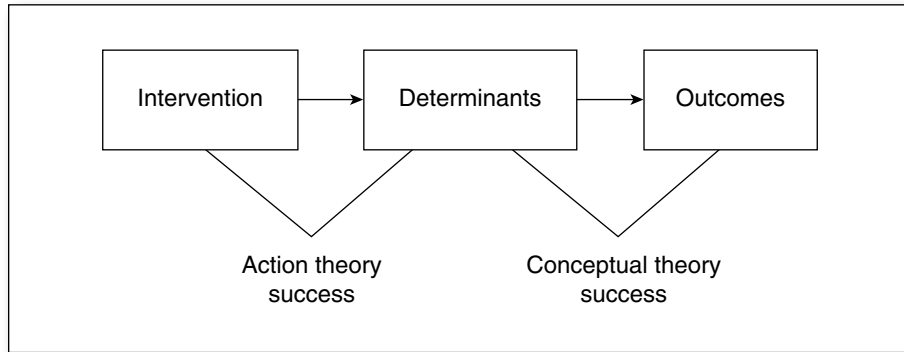
When to Use an Intervening Mechanism Evaluation Approach

Thus far in the history of program development, intervention mechanism evaluation approaches have tended to be applied when there is a need to assess either the accuracy of causal chains or the relative efficacy of intervention components.

Assessing Whether Causal Chains Functioned as Expected. Evaluators can employ intervening mechanism evaluation to see which portion(s) of the causal chains worked as projected and how this has contributed to the success or failure of the program. Graphic representation of this kind of assessment is provided in Figure 12.9, which indicates that an intervening mechanism evaluation sets about testing two theories. The first is the action theory, which is concerned with the intervention's power to affect the determinant; the second is the conceptual theory, which is concerned with the determinant's power to affect the outcome. The overall impact of causal chains depends on the success of both the action theory and the conceptual theory. In the Adolescent Alcohol Prevention Trial (Donaldson et al., 1994), discussed above and illustrated in Figure 12.4, we find an example of an intervention with accurate causal chains. But even when poor action theory or poor conceptual theory is revealed through an intervening mechanism evaluation, stakeholders may benefit from the useful information that the evaluation obtains. Intervening mechanism evaluation is useful in situations such as the following:

- *When action theory fails.* Failed action theory means that an intervention fails to affect its determinant. When this occurs, but the conceptual program theory is sound, then at least the concept that this determinant is a cause of this problem is at least roughly sound. Nevertheless, program designers need to restructure the program intervention to better activate the determinant.

Figure 12.9 Action Theory and Conceptual Theory in the Intervening Mechanism Evaluation



- *When conceptual theory fails.* Failed conceptual theory means that, although an intervention successfully activates its determinant, the determinant then fails to affect the outcome. This evaluation result suggests that program designers have made an invalid assumption about causes of the problem—a fundamental flaw in their model that can be corrected by developing a better conceptual program theory. The above example of the antismoking program evaluated by Chen et al. (1988) illustrates conceptual program theory failure.
- *When action and conceptual theory fail.* Data from an intervening mechanism evaluation that suggest scant association between intervention and determinant, and furthermore between determinant and outcome, indicate that both action theory and conceptual theory are invalid. Continuing a program whose action theory and conceptual theory are both flawed is usually not worthwhile. Heeding the lessons learned from the failure, program designers can begin at the beginning, with a reconceptualization of the problem, the intervention, and the determinant.

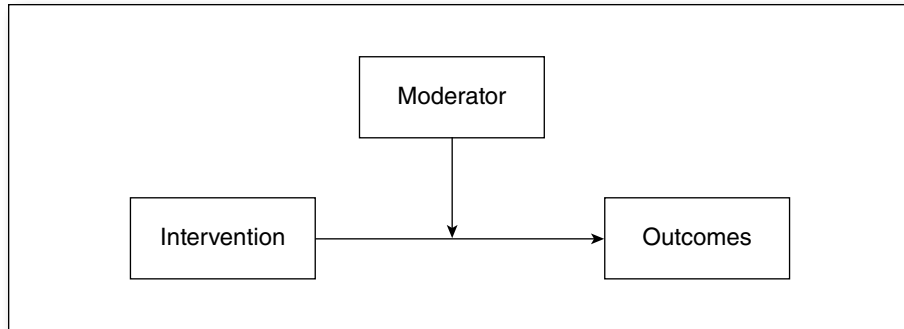
Assessing the Effectiveness of Intervention Components. Another reason to apply intervening mechanism evaluation is to assess the relative effectiveness of an intervention's components. That is, for programs with multiple determinants, a case can be made for which of the determinants the intervention is most effective and for which the intervention should be strengthened. Schneider,

Ituarte, and Stokols (1993) evaluated a program promoting bicycle helmets. The program included a “bicycle rodeo,” physician education, direct-mail communication, coupons for discounted helmet purchase, and telephone communication. Using intervening mechanism evaluation, the researchers determined that the degree of parents’ worry over bicycle accidents is a determinant mediating between the intervention and helmet ownership, an outcome. In addition, analysis of the relationship between intervention components and determinants identified the two intervention components having the greatest influence on children’s helmet ownership: physician advising of need for a helmet and telephone communication. This information suggested that rates of helmet use in a community are most likely to rise when an intervention employs interpersonal education efforts.

Another study by Kristal et al. (2000), discussed earlier in this book, offers a second example of applying intervening mechanism evaluation to compare components’ effectiveness. The research team and stakeholders in a nutrition program had selected three determinants of the outcome, dietary change: predisposing factors, enabling factors, and stage of change. According to data, all three substantially affected dietary change, but only two—predisposing factors and stage of change—were affected by the intervention. The researchers proposed that the enabling factors determinant, the one not responding to the intervention, would be alterable by some as-yet-unknown intervention component, and such a change to the program would enhance the overall intervention effect. In other words, to strengthen this program for future use, activity affecting the enabling factors should be explored, devised, and implemented.

THE MODERATING MECHANISM EVALUATION APPROACH

The second type of theory-driven outcome evaluation is the moderating mechanism evaluation. The moderating mechanism evaluation approach to theory-driven evaluation involves assessing one or more factors in a program’s implementation that condition, or moderate, the intervention’s effect on the outcome. The factors are called *moderators*. Figure 12.10 presents the basic model for the moderating mechanism evaluation. The moderating mechanism is represented by the arrow drawn from each moderator to the midpoint of another arrow that is located between intervention and outcome, delineating the way in which the moderator conditions the intervention-outcome relationship. For example, the effectiveness of family counseling may

Figure 12.10 The Basic Model of Moderating Mechanism Evaluation

depend on the trust maintained between counselor and clients. The level of trust would be a moderator that conditioned the relationship between counseling and the outcome of counseling. Generally speaking, moderators can be clients' sociodemographic characteristics (e.g., race, gender, education, age); implementers' characteristics and styles (e.g., enthusiasm, commitment, skills, race, gender); features of the client-implementer relationship (e.g., trust, compatibility, race, gender); the level of implementation fidelity; and the mode and setting of service delivery (e.g., formal vs. informal, centralized vs. decentralized, rural vs. urban, kind of organizational climate, and integrity of the intervention).

Constructing Moderating Mechanism Evaluation Models

The base of information needed to model a moderating mechanism evaluation can be drawn from stakeholders' ideas and experiences or from the literature. From there on, the strategies used to construct the moderating mechanism model are those discussed above for construction of the intervening mechanism model. Up-to-date moderating mechanism evaluation mainly employs quantitative methods to construct models and analyze data. The equation used for analysis typically includes main effects (for intervention and moderator separately) and an interaction effect (intervention and moderator jointly). If the intervention is conditioned by the moderator, the interaction effect should be statistically significant.

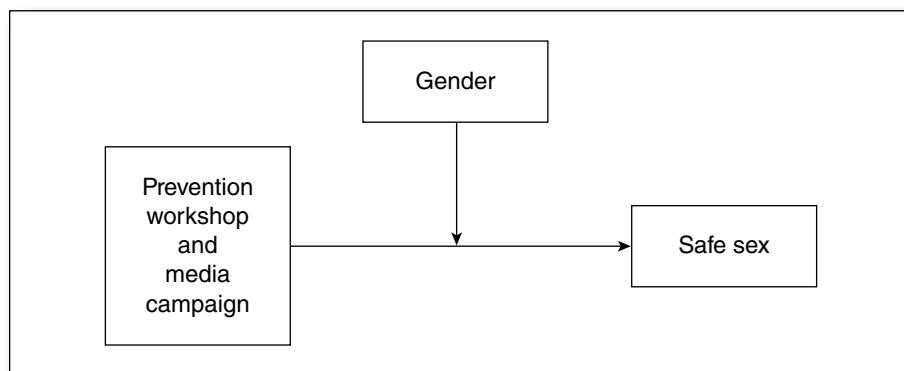
Examples of Moderating Mechanism Evaluation

Two examples have been selected to demonstrate moderating mechanism evaluation: a case in which sociodemographic variables serve as moderators and a case in which the integrity of the implementation serves as the moderator. In the first example, tests were run on a sociodemographic variable, asking whether it conditioned the relationships assumed within the model of an alcohol abuse prevention program (O'Leary, Jemmott, Goodhart, & Gebelt, 1996). The model is illustrated in Figure 12.11. The intervention consisted of workshops in college classrooms and dormitories, as well as exhibits and a print (newspaper) campaign. Outcome measures were the number of sex partners and the total number of unprotected acts of sexual intercourse. The moderator was gender.

Analysis of the data collected concerning this program showed no statistically significant relationship between intervention and outcomes, which, however, was *not* tantamount to finding that no relationship between the two was possible. Indeed, further analysis of moderating processes in the model showed the intervention-outcome relationship was moderated by gender. When all results were in, the evaluation concluded that the intervention had, in fact, reduced risky encounters among males substantially. It was among females that risky behavior remained unaffected.

In another example of moderating mechanism evaluation, we see a valuable demonstration of a program implementation's power to moderate the

Figure 12.11 Social and Demographic Variables as Moderators



SOURCE: Adapted from O'Leary, Jemmott, Goodhart, & Gebelt (1996).

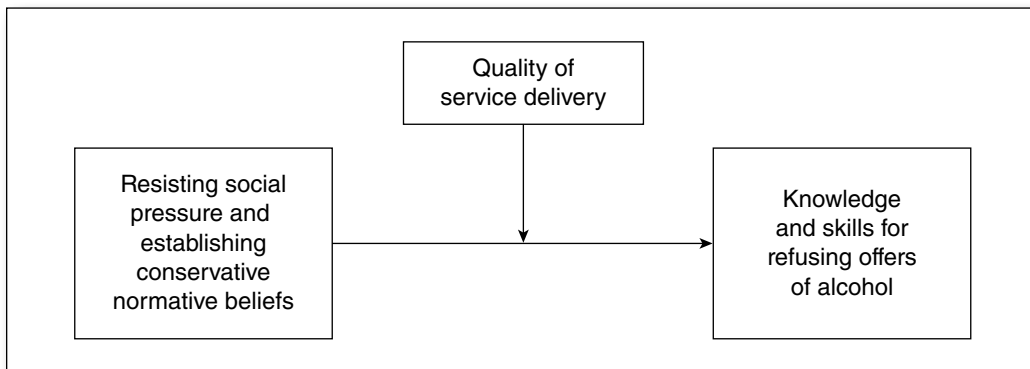
intervention-outcome relationship. This evaluation tested whether quality of service delivery moderated any other program component (Hansen, Graham, Wolkenstein, & Rohrbach, 1991). The intervention involved consisted of curricula intended to train youngsters to resist overt social pressure to use alcohol and other illegal substances, as well as to establish conservative normative beliefs about the social and health consequences of substance use. The outcomes to be evaluated were seven indicators that pertained either to knowledge and perceptions of social pressure to use substances or to related refusal skills. Measures of quality of service delivery were ratings assigned by the program's expert stakeholders, along with a trained observer's ratings on items including enthusiasm, student responsiveness, student participation, classroom control, interaction with students, and meeting program goals. Figure 12.12 presents the model of this fairly complex program.

The evaluation suggested that intervention activities did affect some outcome measures. Just as importantly, it showed that the quality of service delivery had significantly moderated relationships between the intervention and three of the seven anticipated outcomes. These results led to a conclusion that the intervention worked best with high-quality service delivery.

Advanced Moderating Mechanism Models

In some cases, a moderating mechanism and an intervening mechanism are at work simultaneously in a program. Donaldson (2001) indicated that a moderator

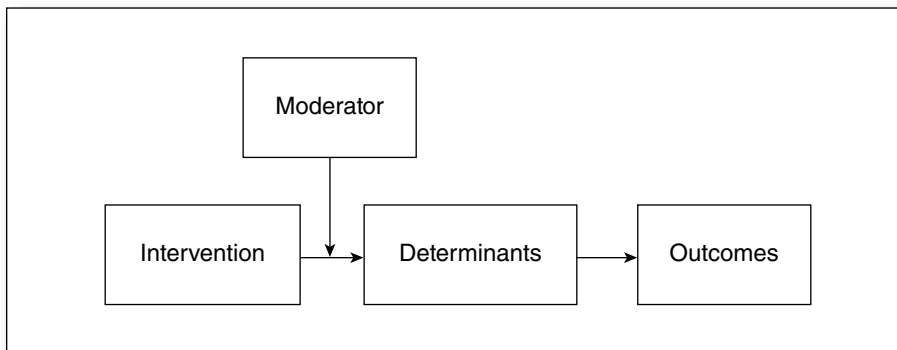
Figure 12.12 Service Delivery as a Moderator



SOURCE: Adapted from Hansen, Graham, Wolkenstein, & Rohrbach (1991).

may condition the relationship between intervention and determinant or between determinant and outcome. The first possibility is illustrated in Figure 12.13. The model in this figure shows the effect of an intervention on a determinant being conditioned by a moderator, as in the earlier example of the AIDS 101 intervention. In that example, the determinant supporting the intervention was increased HIV-prevention knowledge and skills.

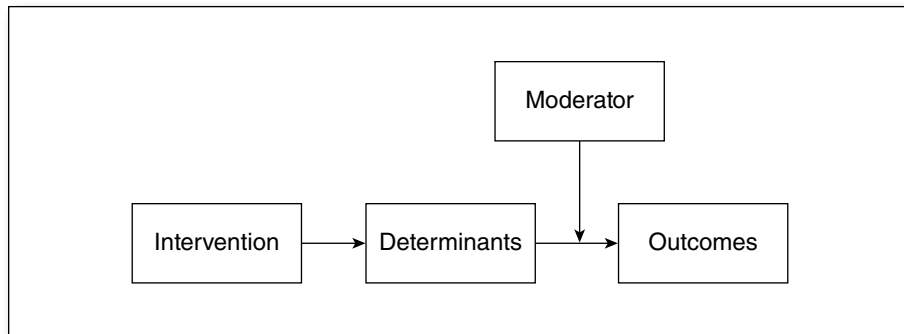
Figure 12.13 Moderating Mechanism Conditions the Relationship Between Intervention and Determinant



The second possibility, presented in Figure 12.14, illustrates a further example. This model shows how a moderator that influences the relationship between a determinant and an outcome—for instance, the clients’ education levels—might be involved in the success of AIDS 101. An education moderator may reduce the usefulness of AIDS 101 for highly educated clients, possibly because their existing knowledge and skills are sufficient for avoiding HIV. On the other hand, moderation by education level may allow AIDS 101 to be very useful for less educated clients, as they may well encounter little medical information in their daily lives.

Another example that illustrates the second possibility is a condom distribution program. For a condom distribution program, the determinant might be availability of a condom. The model in Figure 12.14 demonstrates how condom availability’s power to promote safe sex may be conditioned on sexual partners’ acquiescence to values endorsing male dominance. Thus, even with condoms readily available, partners (especially men) who assume men’s preferences should predominate are less likely than other individuals to use a condom

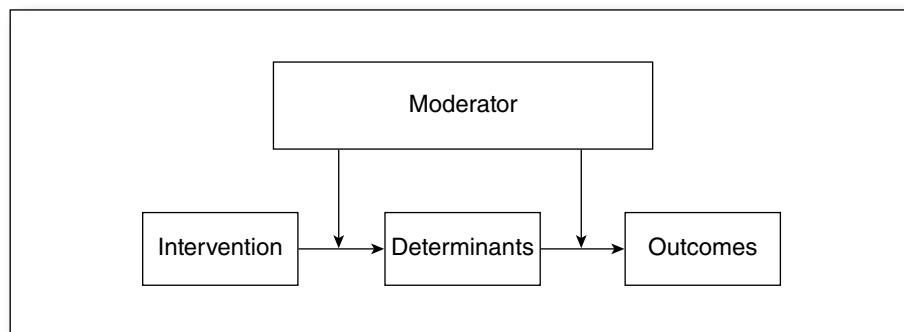
Figure 12.14 Moderating Mechanism Conditions the Relationship Between Determinant and Outcome



for safe sex. On the other hand, the model suggests, partners valuing gender equality will perhaps use condoms if they are available.

In addition, there is another possibility: A moderator may condition at once the relationship of the intervention to the determinant and the relationship of the determinant to the outcome. The case is illustrated in Figure 12.15. Dual conditioning of this type would be noted in a condom distribution program in which a value placed on male dominance moderated partners' willingness to receive and carry condoms, as well as their actual condom-using behavior.

Figure 12.15 Moderating Mechanism Conditions Relationships Among Intervention, Determinant, and Outcome



When to Use a Moderating Mechanism Evaluation Approach

The moderating mechanism evaluation approach may be utilized when the purpose of evaluation is to (a) identify a need to tailor interventions for different groups or (b) compare the relative effectiveness of a program's structural options. Determining a need to tailor interventions belies the widespread assumption that an intervention works or does not work equally for all kinds of target group members. The truth is that a target group's membership often comprises a variety of cultural, social, and economic backgrounds. Thus, interventions should be *expected* to work better for some members than others. The O'Leary research team's 1996 evaluation of the alcohol-abuse/safe-sex campaign showed that a moderating mechanism evaluation is equipped to comment on an intervention's generalizability across a target population. Moderating mechanism evaluation, in other words, discloses which intervention works for whom. Should strong differential effects of the intervention be deemed an issue within a target group, it becomes important to tailor versions of the intervention to the differential needs of subgroups within the targeted population.

As for the second task with which moderating mechanism evaluation can be charged, we must recognize the consistent truth that a program could often be structured in various ways. When more than one option presents itself, a moderating mechanism evaluation allows evaluators to formally test whether one option will be more effective than another. For example, stakeholders may wonder whether a particular intervention is especially effective if clients and implementers share a racial/ethnic background. Or, stakeholders may ask whether an intervention's efficacy depends on the delivery setting (e.g., home vs. workplace) or delivery mode (group vs. individual), and so on. A moderating mechanism evaluation approach is well-suited for assessing interaction between the intervention and other program elements.

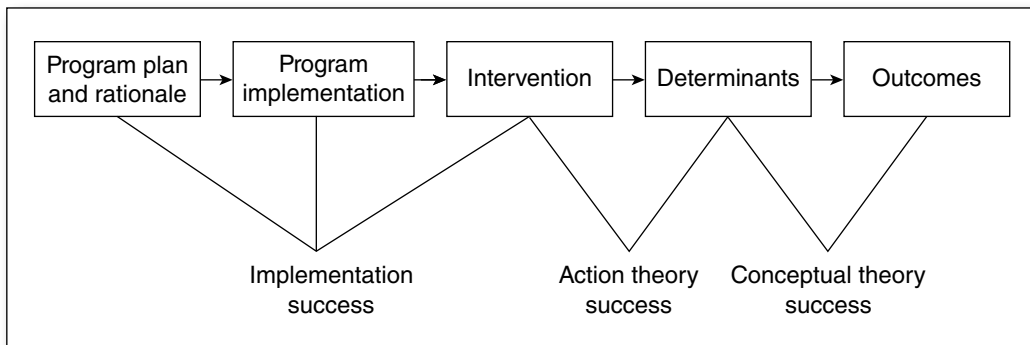
THE INTEGRATIVE PROCESS/OUTCOME EVALUATION APPROACH

The third and final theory-driven outcome evaluation included in this book is the integrative process/outcome evaluation. This type of evaluation involves the systematic assessment of (a) the crucial assumptions beneath implementation and (b) the causal processes of a program. In other words, integrative process/outcome evaluation weighs the range of elements in a program theory

(see Figure 12.2). This consummately comprehensive assessment provides a network of information about what works and what does not work in a program, from implementation processes to causal processes to effects on outcomes. Such a thorough analysis of potential pathways enlightens stakeholders as to how their program truly operates, providing the knowledge they will need to meet the accountability and program improvement requirements they face. Because this kind of comprehensive evaluation integrates both process and outcome evaluation, its relationship with other parts of a program can be illustrated as in Figure 12.16.

In Figure 12.16, one sees that an integrative process/outcome evaluation systematically assesses the implementation of the program plan and the truthfulness of the program rationale. One also sees what a challenge it can be to design a successful intervention program. Successfully implementing the program design in the field is also complex. The model makes clear that “implementation success” occurs only when an intervention appropriately activates a change process. Implementation success, then, is vital to the entire change process: If implementation fails, everything fails. Nevertheless, a successful implementation *guarantees* nothing about program success. In other words, it is necessary but not sufficient for program success. If a program is to be effective, its action theory and its conceptual program theory must succeed along with its implementation. Invalidity of either action theory or conceptual program theory could spell its doom. Comprehensive, systematic integrative process/outcome evaluation abundantly fleshes out these assumed underlying mechanisms. It thus provides stakeholders virtually any information they require for improving their programs.

Figure 12.16 Linkages of Major Components in an Integrative Process/Outcome Evaluation



Research Methods and Strategies Associated With Integrative Process/Outcome Evaluation

Integrative process/outcome evaluation often relies on mixed methods of collecting qualitative and quantitative data. Qualitative methods are preferred for formulating a model and acquiring data to test it, especially if the data describe some aspect of the implementation process. Designing an integrative process/outcome evaluation from start to finish involves the following general steps:

1. *Clarify the program theory.* Evaluators work with key stakeholders, including program designers, to clarify the program theory. This step incorporates both specifying the change model, or program scope (see Chapter 4), and specifying the action model, or action plan (see Chapter 5).

2. *Collect and analyze data.* Guided by the program theory, empirical data are collected that will demonstrate how the theory operates in the field. Evaluators usually need to use mixed methods (qualitative and quantitative) to collect data pertinent to both the action model and the change model. Qualitative methods, however, are particularly useful for probing the reasons why a component is not working as well as expected. Data concerning linkages among program elements, from implementation to causal processes to final outcomes, need to be analyzed.

3. *Characterize the program in its entirety, then by its parts.* When all data have been analyzed and compiled, the evaluator provides to stakeholders a written assessment of whether the program as a whole is effectively reaching its goals. Incorporated in the report should be detailed analyses of the important parts of the program as well, covering how well each is working and how each contributes to or hinders the achievement of program goals.

Examples of Integrative Process/Outcome Evaluation

Some readers may recognize that the garbage reduction program discussed in Chapter 5 is an example of integrative process/outcome evaluation. This section will turn to two other examples to illustrate the benefits of integrative process/outcome evaluation and to model some of the strategies involved.

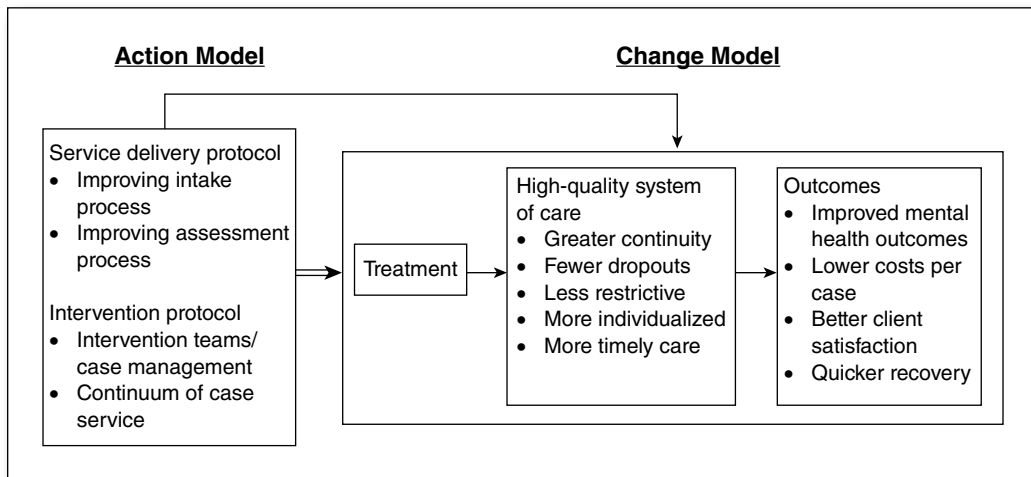
Fort Bragg Child and Adolescent Mental Health Demonstration

The mental health services staff at the US Army's Fort Bragg base in North Carolina assessed whether the continuum of services provided to children

actually improved treatment outcomes and reduced costs of care per client, as program goals required (Bickman, 1996). The program provided traditional mental health services, such as outpatient therapy and acute inpatient care, and more innovative services, such as case management, in-home therapy, after-school group treatment services, therapeutic homes, and 24-hour crisis management teams. Fort Bragg families who requested services underwent comprehensive intake and assessment. In an attempt to control costs, the staff offered a continuum of services, at least some of which were expected to be appropriate for any given child. Evaluators used interviews, document review, and focus group meetings to develop the program theory for the project, which is delineated in Figure 12.17.

The action model seen in the figure shows that the Fort Bragg Child and Adolescent Mental Health Demonstration emphasized the treatment and service delivery protocols. The change model posits as a determinant a better system of care, which mediates between intervention and outcome. Evaluation data suggested that the Fort Bragg demonstration had improved both access to care and quality of care, but clinical outcomes were disappointing. Little difference appeared between the demonstration program and comparison programs when it came to improved mental health, lower costs per client, quicker

Figure 12.17 Program Theory of the Child and Adult Mental Health Demonstration



SOURCE: Adapted from Bickman (1996).

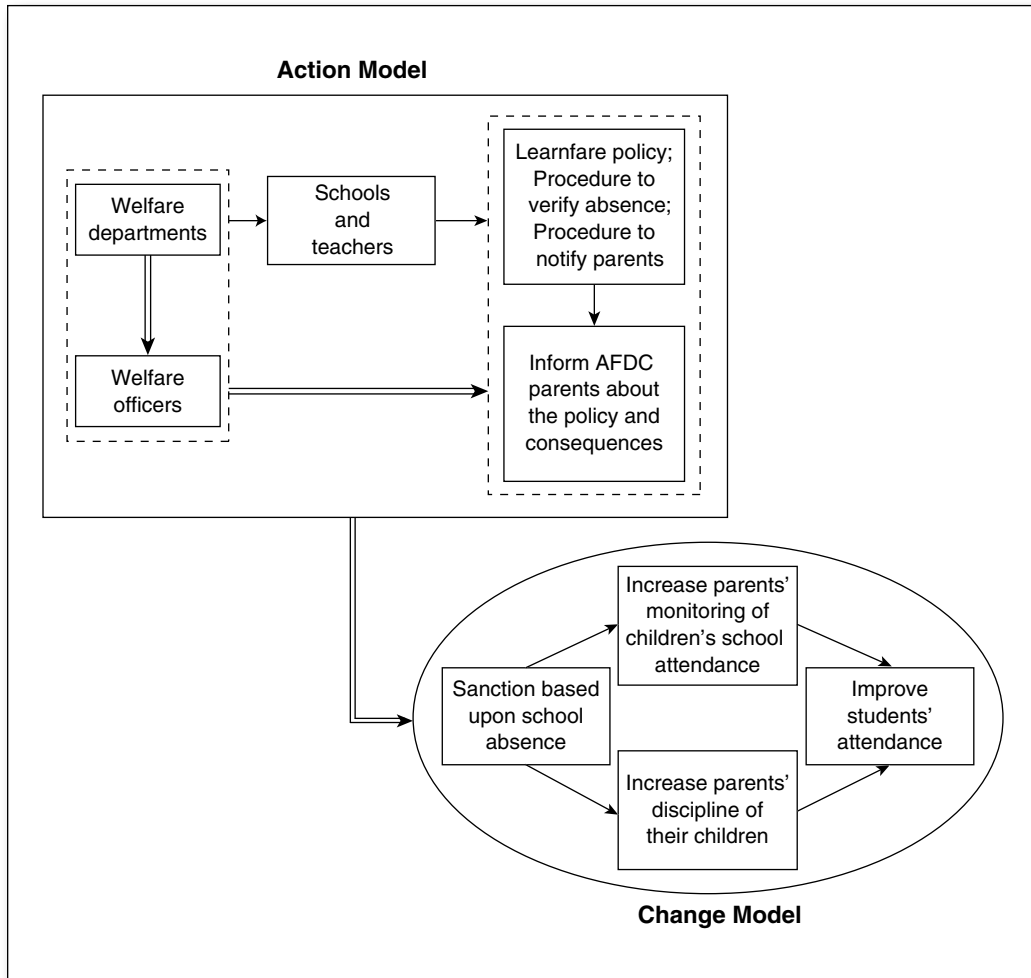
recovery, and client satisfaction. Thus, the demonstration program had implementation success and action theory success but experienced a failure of its conceptual program theory.

Learnfare

A second example of integrative process/outcome evaluation is recorded in Ethridge and Percy's (1993) evaluation of a Wisconsin welfare program called Learnfare. Learnfare was, at heart, a welfare policy tying Aid to Families with Dependent Children (AFDC) payments to the school attendance of recipients' children. Specifically, children in families receiving AFDC payments were monitored by Learnfare to see how regularly they attended school. Learnfare sanctioned the family of a child who recorded two or more unexcused absences in a given month by cutting its monthly AFDC payment by the portion normally earmarked for that dependent child. The aim was to increase parents' involvement in schooling of their children, especially in terms of enhancing the educational achievement of teenagers, in an effort to decrease the probability that the children would depend on AFDC when they grew up. Ethridge and Percy used program theory to understand how the policy was being implemented and its limitations. The program theory underlying Learnfare is illustrated in Figure 12.18.

Figure 12.18 indicates that Ethridge and Percy identified two determinants that were crucial to the effectiveness of the Learnfare policy's change model: (a) an increase in parents' responsiveness to attendance-related communications from schools, combined with an increase in parents' own monitoring of children's school attendance, and (b) increased capacity of parents to control school attendance-related behaviors of children. The evaluators questioned the validity of the first determinant. Low rates of literacy are common among parents receiving AFDC, they pointed out; it is not farfetched to propose that many such parents are unable to read and understand communications from schools concerning attendance. The authors also felt that there were problems with the second determinant stakeholders had specified. They suggested that because many families are single-parent households, impoverished and susceptible to various problems such as frequent illness and unstable employment, expecting parents to routinely monitor and control children's behavior might be unrealistic. Thus, the change model on which Learnfare was founded could not, they believed, operate as stakeholders anticipated.

Learnfare's action model also exhibited major weaknesses, according to Ethridge and Percy (1993), especially in its target group, implementing organizations, and relationships with peer organizations. Some of the problems that the evaluators brought up seem unmistakable in hindsight:

Figure 12.18 Program Theory Underlying Learnfare

SOURCE: Adapted from Ethridge & Percy (1993).

- “*Unexcused absence*” was not clearly defined. With unexcused absences triggering sanctions, implementers, parents, and students needed a precise explanation of what unexcused absences were. Learnfare had given no definition of an unexcused absence that would make its attendance requirements comprehensible and thus make the program workable. Schools’ individual

policies about acceptable absences showed greatest variation when it came to family vacations, hunting or fishing trips, and family business activities.

- *Gathering attendance information was difficult.* Learnfare depended on the transmittal of attendance information from schools to the welfare system. This sounds straightforward until one realizes that state law mandates the privacy of school records. Unauthorized sharing of attendance data between schools and county or state welfare systems violated the law, so each AFDC parent needed to waive this right to privacy and direct school officials to reveal the data. This seriously delayed transmission of attendance data. Furthermore, for Learnfare to succeed, schools needed to transmit accurate data in a timely way, and this proved problematic. Schools had difficulty keeping their attendance data current. County welfare officials needed to act on data that often arrived late, was hard to interpret, and was filtered through the policies and systems of many idiosyncratic school districts. Another law required schools to follow elaborate verification procedures to increase the accuracy of attendance reports. In the wake of all this, 3 or 4 months could pass between the unexcused absence and the punitive loss of funds.

- *Communication with AFDC recipients was complicated.* If Learnfare were to work, AFDC recipients had to be aware of the sanctions and the reasons for them. They also needed to be notified when a child's truancy had triggered a benefit reduction. A substantial effort was made to disseminate information about Learnfare to the target group. However, the program persistently did not give parents sufficient notification about specific dates of attendance violations. The program had difficulty communicating to parents the dates of first and second (and subsequent) unexcused absences. This failure was the result of an unproductive relationship between welfare officials and local school officials. School officials had not been invited to help plan for Learnfare and its attendant impacts on educational objectives, school support personnel, a school's own attendance recording and monitoring procedures, and nearby alternative schools.

THEORY-DRIVEN OUTCOME EVALUATION AND UNINTENDED EFFECTS

In evaluation parlance, *unintended effects* are the effects made by a program that are outside the scope of program goals. The tendency for social action to generate unintended effects has long been articulated in the literature (e.g., Weber, 1947). However, when enlightenment is a central purpose of evaluation,

it is possible for the evaluator to use evaluation design to manage the potential for unintended effects. Unintended effects can be either positive or negative. Incarcerating juvenile delinquents perhaps has the negative effect of allowing them to learn criminal skills from veteran offenders. The 55-mile-per-hour speed limit, in contrast, was intended to create a fuel conservation effect and then was found to also boost highway safety, a positive unintended effect (Clotfelter & Hahn, 1978). A great advantage of theory-driven outcome evaluation is the detection of unintended effects that it facilitates. Unintended effects are detected during theory-driven outcome evaluation by two general strategies: formal specification and field study.

Formal Specification of Possible Unintended Effects

A strategy to specify formally any possible unintended effects involves rendering each such effect as a hypothesis that the evaluation can test. The evaluator has two options here: to use existing work to infer unintended effects or to ask stakeholders to brainstorm possible unintended effects. The former option tasks evaluators with reviewing existing theories and studies that are relevant to the program. This permits the formulation of hypotheses about possible unintended effects of the current program, which can then be tested. For example, in evaluating compulsory seat belt legislation and its effect on vehicle occupant casualties, Conybeare (1980) harnessed existing theory and knowledge and posited the potential unintended effect of a rise in the casualty rate for nonoccupants. Conybeare's reasoning was that drivers experiencing the legislation-induced security of being strapped in might begin to drive less attentively, resulting in more accidents with pedestrians, bicyclists, and so on. Such a hypothesis was included in the evaluation design. The data confirmed that the legislation had created both the stakeholders' intended and the evaluator's unintended effect.

The second option for detecting unintended effects during evaluation is to prompt stakeholders to brainstorm possible unintended effects. Implementers and other stakeholders have a great deal of experience in working with clients, communities, and implementations. They are a good source of ideas about unintended effects worthy of inclusion in an evaluation. For example, I participated in evaluating an after-school project designed to offer academic tutoring, recreation activities, and drug use prevention education to adolescents of rival South Asian ethnic backgrounds. The program goals were to enhance the target group's success in school and cultivate in them attitudes and skills for

resisting drug use. As the program theory was being formulated, counselors with the program pointed out an important potential unintended effect, which was that the interaction fostered by program activities might weaken the participants' prejudices toward Asian ethnic groups not their own. In many Asian communities, including the one that was home to this project, interethnic group prejudices are a pressing concern, as they often lead to conflict and even violence. The hypothesis suggested by the program counselors was included in the evaluation design.

Field Study Detection of Unintended Effects of Implementation

The integrative process/outcome evaluation usually combines both qualitative and quantitative methods in order to intensively examine an implementation process. The evaluator in the field will benefit from understanding how a program is being implemented, how its clients and implementers are responding to it, and how it is interacting with its environment. Field observation provides a good opportunity for evaluators to identify and investigate potential unintended effects. For example, I had reason to visit the site where a successful alcohol treatment program was being implemented within an institution that confined its residents. The quality of implementation was very impressive. But it was also clear that many clients smoked cigarettes during breaks and after meals. Perhaps smoking relieved the stress of eschewing alcohol. Staff acknowledged that smoking by clients was permitted. This field observation led to a hypothesis about an unintended effect of the program: Clients who cease consuming alcohol may begin to consume tobacco products in greater quantities.

The assessment by Gottlieb et al. (1992) of the effects of a workplace smoking policy (used as an example in Chapter 7) also provides a good illustration of the detection of unintended effects during implementation. In interviewing smokers and nonsmokers employed by the affected organization, the study found four unintended effects:

- *Diminished air quality in designated smoking areas.* Smokers were relegated to certain smoking areas, where they tended to congregate, filling the air with the smoke of many cigarettes rather than one. Air quality was considerably diminished in the smoking areas, especially poorly ventilated ones. The odor of cigarette smoke became so thick that it even reached nonsmokers who were at some distance from the smoking areas.

- *Disruption of work.* Implementing the smoking policy led to the disruption of work. Clerical staff left their desks more often; nonclerical employees began to move their work with them to smoking areas and the cafeteria.
- *Disruption of communication.* Once the policy was in place, nonsmokers were less likely than before to discuss smoking-related issues with smokers; the percentage of nonsmokers requesting a co-worker not to smoke near them decreased. Some employees expressed concern that the smoking policy polarized the smoking and nonsmoking employees.
- *Purported creation of “deviance.”* After the policy had been implemented, employees surfaced who said they were concerned that tobacco use was being demonized, resulting in smokers being discriminated against more noticeably. Smokers were, they said, increasingly viewed as deviants.

The use of both quantitative and qualitative methods during integrative process/outcome evaluation opens an opportunity for evaluators to appraise a program's unintended effects. In the Chapter 7 example of the anti-drug program launched by Taiwan's Ministry of Education (Chen, 1997), quantitative outcome data argued that the program was a strong success. However, some substantive problems identified in the implementation process cast doubt on that conclusion, in spite of the quantitative data. Evaluators returned to the field to interview key informants and probe them about these problems. They learned that in the local schools, staff feared the Ministry of Education would use the data to punish administrators and teachers at schools with more drug-using students. They sought to protect themselves by reporting fewer cases than they knew to exist. Thus, the outcome data they provided were not a valid reflection of the program's effect.

A REPLY TO CRITICISMS OF THEORY-DRIVEN OUTCOME EVALUATION

In spite of the popularity of theory-driven outcome evaluation (Coryn et al., 2011), earlier literature (Scriven, 1998; Stufflebeam, 2001) claimed to have identified certain flaws in it. Although recent literature (Chen & Turner, 2012; Coryn et al., 2011; Hansen & Vedung, 2010) shows that evaluators increasingly recognize the usefulness of stakeholder theory and the theory-driven evaluation, it is well worth briefly considering some of the criticisms, as well as responses that address them.

To begin, Scriven (1998) professed that the task of program evaluation is to assess the merit or worthiness of a program. Such an assessment is indistinguishable from an evaluation of products such as dishwashers and automobiles. Just

as in the work of *Consumer Reports*, the credibility of program evaluation rests on evaluators' objectivity. Because there is no need to understand how and why a dishwasher works in order to assess its merit, there is no need, either, to understand how and why a program works in order to assess its merit. This book has, from the start, established a different viewpoint. What a program evaluation is and what a program evaluator does far exceed the limited role assigned by Scriven. The whole of the reasoning supporting this argument is available in Chapter 1, but its larger points are summarized here.

Scriven ignored the human factors in program evaluation. Product evaluators do not need to be concerned with a dishwasher's or automobile's views. However, program evaluators have to consider stakeholders' views and concerns for an evaluation to be considered credible and fair. Scriven's model narrowly focuses on assessing the intrinsic merit of a program; for example, it may be justifiable to assess a television based mainly upon its intrinsic features, such as the quality of picture, sound, durability, style, and so on. However, an evaluation of a program needs to consider both intrinsic and extrinsic factors. Because human beings highly react to both program participation and evaluation, the merit of a program cannot be assessed adequately without an awareness of how the accomplishment of goals has been pursued. Scriven's model for program evaluation could therefore be misleading. Furthermore, in contrast to product evaluation, program evaluation commonly has as a central purpose program improvement, and Scriven's limited view of program evaluation does not correspond to this reality. As discussed in Chapter 1, an effective practical evaluation must be future action directed, have both scientific and stakeholder credibility, and take a holistic approach. The theory-driven outcome evaluation model can supply these qualities.

Stufflebeam (2001) also did not recommend theory-driven outcome evaluation. However, his primary criticisms inaccurately reflect the nature and procedures of theory-driven outcome evaluation. He alleged the following:

1. Evaluators using program theory might, in fact, develop their own theory and evaluate it, creating a conflict of interest.
2. Theory-driven outcome evaluation depends on a base of sound theory, which few programs can point to as their foundation.
3. Evaluators using theory-driven outcome evaluation tend to displace whatever program staff members have been using to create the program design.
4. Theory-based outcome evaluation is too difficult to do correctly.

Stufflebeam's criticisms plainly misconstrue authentic program theory. For example, concerning his first objection, a program theory is always the *stakeholders'* program theory. Stakeholders have made crucial assumptions, and the evaluator's role is to make those assumptions explicit to facilitate assessment of the stakeholders' program. To my knowledge, no one advocates for the imposition of evaluators' own agendas on stakeholders' programs, which they then evaluate. Nor are there examples of evaluations that have succumbed to such a practice.

Stufflebeam's second criticism was addressed in Chapter 3, where I explained that a program to be evaluated can be based on explicit, validated scientific theory *or* on stakeholders' implicit theory. In either case, theory-driven outcome evaluation is applicable, and its success will be entirely unrelated to the philosophical source of the program. Theory-driven outcome evaluation builds mainly upon evaluators' ability to help stakeholders document an explicit program theory and systematically implement it.

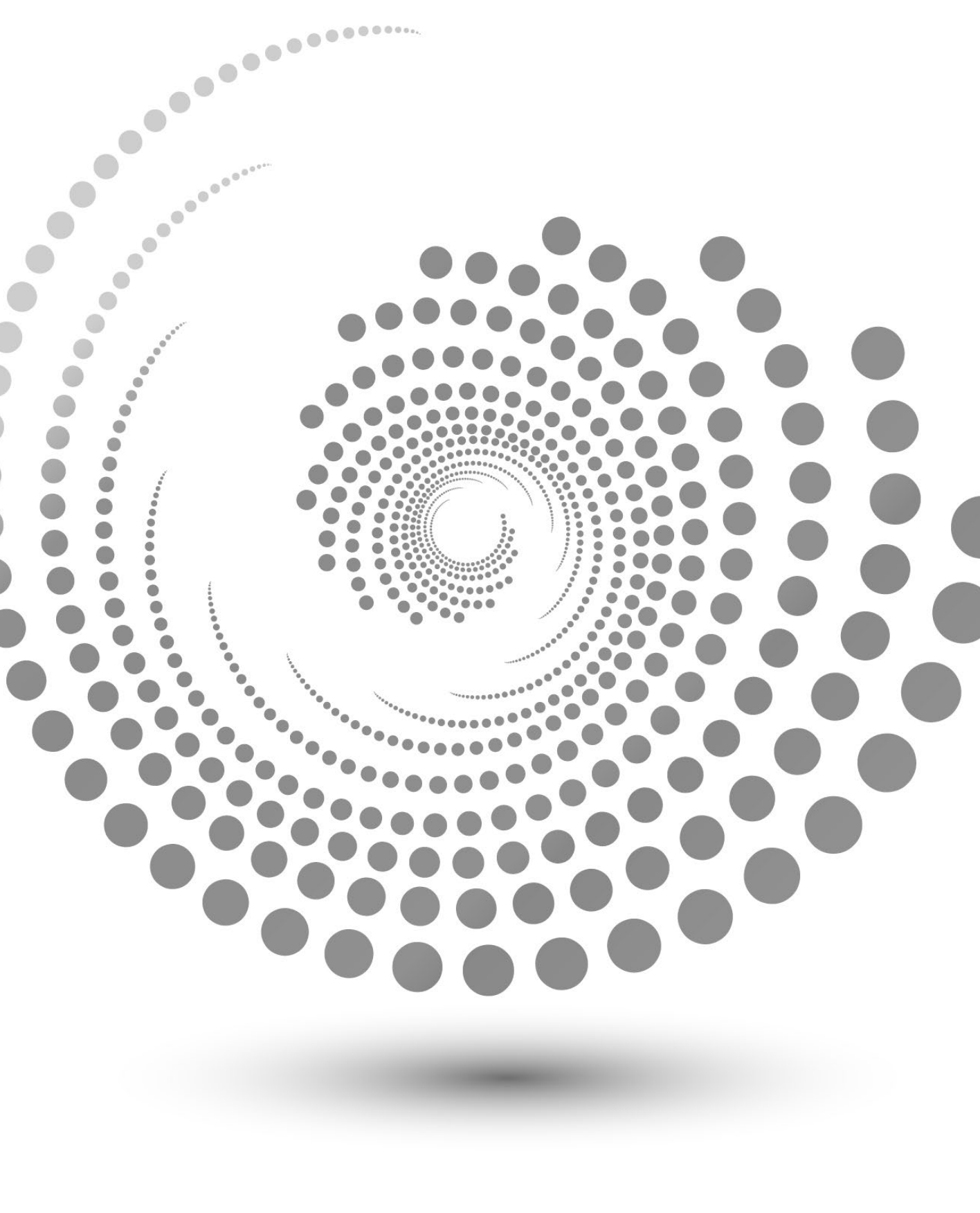
Stufflebeam's third criticism does raise one point not to be disputed: Program design is the program staff's job. However, assisting or facilitating does not equate with usurping, and professional program evaluators are well aware of this. Thus, through theory-driven outcome evaluation, stakeholders are motivated to design and implement programs better and more thoroughly using the comprehensive, meaningful information made available in the evaluation's results.

Finally, Stufflebeam's stance on the difficulty of conducting theory-driven outcome evaluation correctly has not been substantiated with any evidence. On the contrary, there has recently been a growing appreciation among evaluators for the merits of theory-driven evaluation (Chen, 2012a, 2012b; Chen & Turner, 2012; Coryn et al., 2011; Funnell & Rogers, 2011; Hansen & Vedung, 2010). As demonstrated in this chapter, the application of theory-driven outcome evaluation approaches, citing examples in which, through correct application of program theory, the associated procedures were carried out without great difficulty, resulting in very useful conclusions.

QUESTIONS FOR REFLECTION

1. What is theory-driven evaluation? Why is a theory-driven outcome evaluation a hybrid evaluation that serves both accountability and program improvement needs?
2. What is stakeholders' implicit theory? Why is it important?

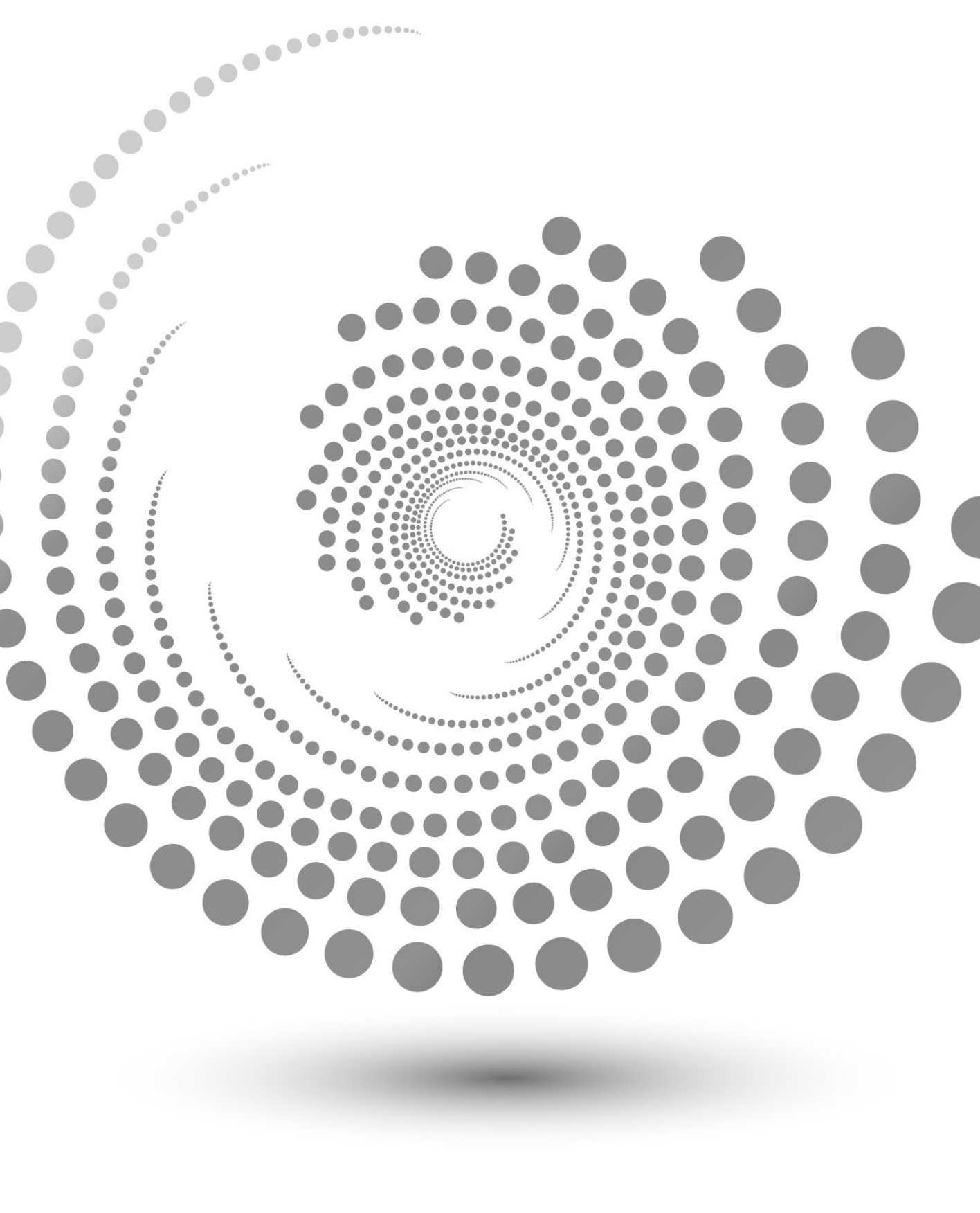
3. Describe some of the issues experienced by stakeholders when trying to develop a program theory.
4. List and briefly explain the guidelines for conducting theory-driven outcome evaluation.
5. Explain the core of theory-driven outcome evaluation. How do all the elements work together to achieve the desired outcome?
6. Define the intervening mechanism evaluation approach. How is it similar to or different from the moderating mechanism evaluation approach? Give examples of intervening mechanisms and moderating mechanisms.
7. Explain the three types of theory-driven outcome evaluations (intervening mechanism, moderating mechanism, and integrative process/outcome evaluation). What components are different, and what components are similar? How does a program evaluator determine the best evaluation to employ?
8. Give real-world examples of the three types of theory-driven outcome evaluations.
9. What are the differences among the following concepts: action theory failure, conceptual theory failure, and implementation failure. Give examples of each type of failure.
10. How do theory-driven outcome evaluations address unintended intervention effects?



PART V

Advanced Issues in Program Evaluation

This part contains three chapters discussing cutting-edge issues in program evaluation. Chapter 13 discusses two studies that illustrate how to proceed when an application of logic models is not as useful as expected. Chapter 14 discusses the relative strengths and limitations of intervention programs as described by formal theory and stakeholder theory. Chapter 15 proposes the bottom-up approach as an alternative to the conventional top-down approach to address issues with evaluation and dissemination.



CHAPTER 13



What to Do if Your Logic Model Does Not Work as Well as Expected

As discussed in Chapter 3, logic models are a popular tool for program planning and evaluation. Many funding agencies in the human services, public health, and social science arenas require applicants to include a logic model in grant applications. Logic models are frequently found to be beneficial. For example, the author and colleagues found the conceptual framework of logic models to be beneficial in the Academic Health Department project (Turner et al., 2014), as it helped evaluators and stakeholders navigate complicated program activities and identify major program components to guide program planning and evaluation.

However, intervention programs vary greatly in terms of their nature, structure, strategies, and purposes. It is therefore likely that logic models work well for many programs, but not for all of them.* Since the format of logic models does not require a detailed specification of contextual factors and causal mechanisms as discussed in Chapter 3, a straightforward application of logic models is likely to confront difficulties or create confusion when programs put a major emphasis on these areas. To help readers understand the issues, this chapter uses two real-world cases to empirically illustrate the nature of the problems. The first of these two studies is a diversity enhancement project; the second, a community health initiative. The chapter will discuss how the problems confronted in

*Several publications make important contributions in pointing out the limitations of logic models. Some also offer solutions for addressing the problems. However, the discussions often remain at a theoretical level without providing real-world examples for illustrating the problems or solutions. For example, Reeler (2007) and Sabatier (1999) criticize logic models on the basis that these models typically prescribe a linear, one-way relationship between one component and another. They argue that, in reality, a program's logic and progression are often nonlinear, moving back and forth among

these two studies were resolved by using the action model/change model schema. The remainder of the chapter will use the lessons learned from these two case studies as a basis to propose strategies for evaluators to more competently and productively apply logic models for describing programs.

A DIVERSITY ENHANCEMENT PROJECT

Description of the Project

This diversity enhancement project focused on increasing the recruitment, retention, and graduation rates of students from diverse and disadvantaged backgrounds, ultimately increasing diversity of the nursing workforce. It sought to achieve this goal through recruitment, retention, and graduation interventions that empowered each student in a nursing college. The program addressed academic, social, physical, financial, professional, and individual needs of the participating students and the social, environmental, and structural determinants that impacted their success.

The project sought to advance the diversity of the nursing workforce by way of the following strategies:

- Incorporating the SAFER Model (Social support, Academic support, Financial support, Empowerment, and Responsibility; Swinney & Dobal, 2008) to address the social, academic, and environmental barriers to success for both undergraduate and graduate students.
- Partnering with community organizations to identify potential students from diverse or disadvantaged backgrounds who were interested in pursuing graduate education.
- Collaborating with inter-professional partners to increase capacity and expertise relative to health equity at the college and in the community.
- Concurrently focusing on the structural, cultural, and environmental milieu within the college to begin a dialogue and change the conversation about diversity, cultural competence, and health disparities.

components. However, they do not provide empirical examples to illustrate these limitations, nor do they propose concrete strategies to address the problems. Renger and Titcomb (2002) point out the importance of clarifying a program's underlying rationale before applying a logic model to describe the program. However, they do not provide real-world examples for illustrating either the problems that occur when underlying rationales are not clarified or the benefits gained with such a clarification.

Applying the Logic Model

Stakeholders invited an evaluation team to assist them with developing a logic model for the project, which is illustrated in Figure 13.1.

As indicated in Figure 13.1, the activity component summarizes the actions that would be taken. It includes the following elements: creating a research team with a governance structure for coordination; partnering with community-based organizations for recruitment; developing a diversity-enhanced curriculum; and providing financial, tutoring, and other services to diverse/disadvantaged students. Through these actions, the program intended to initiate a culture change in the college and achieve goals such as an increase in the enrollment, retention, and graduation rates of diverse/disadvantaged students.

Stakeholders said that the logic model was useful in the sense that it visually represented the major components of the project and their relationships. The model would facilitate communication about the project. However, stakeholders also indicated that the activities component did not sufficiently highlight the environmental changes the project sought to implement. The evaluation team tried to revise Figure 13.1 by grouping elements of the activities component into categories. However, the revised version still did not meet stakeholders' needs. The evaluation team suggested that the action model/change model schema, as discussed in Chapter 3, might be useful because it addresses contextual factors. The stakeholders agreed to give it a try.

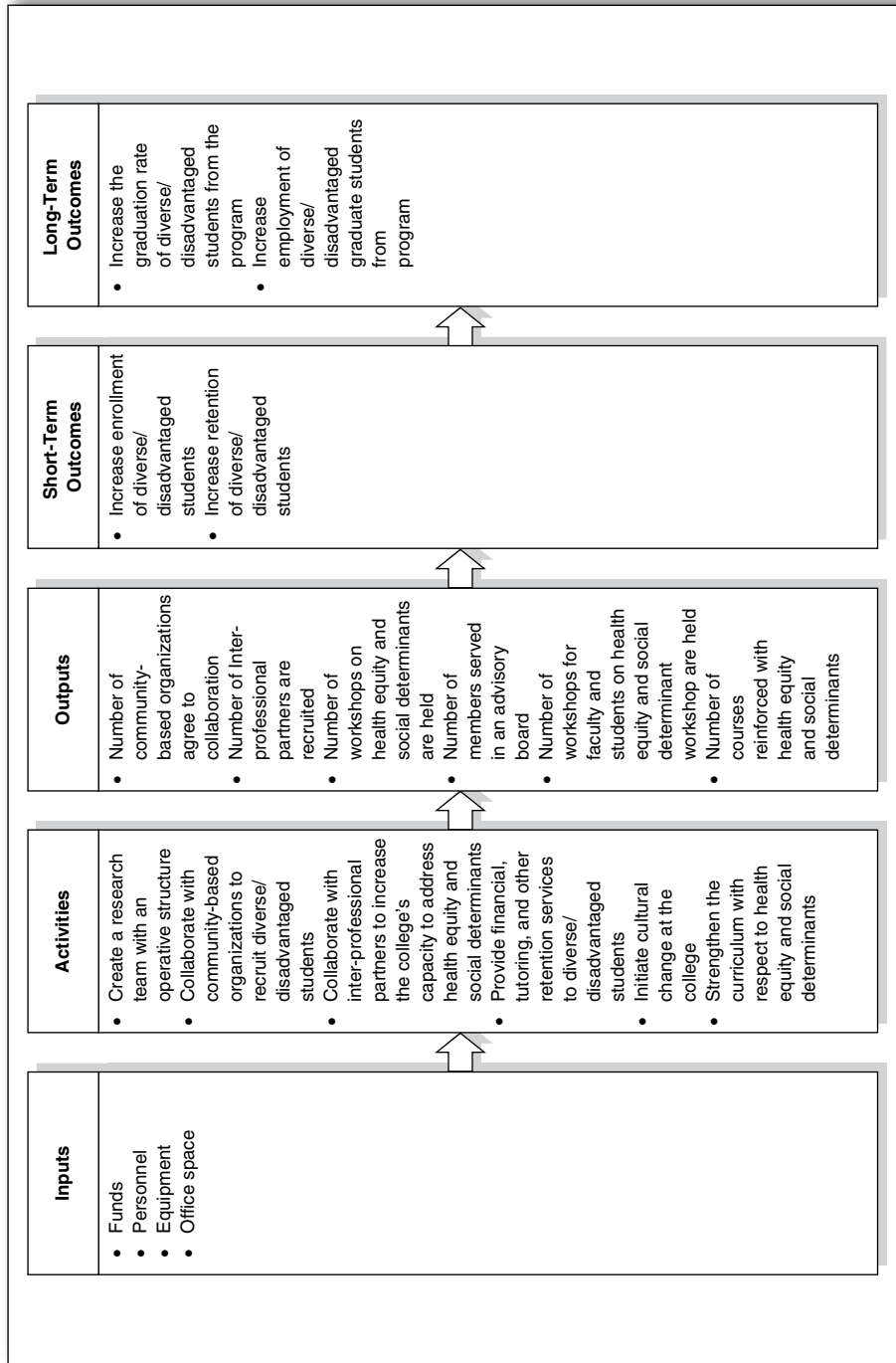
Applying the Action Model/Change Model Schema

The evaluation team introduced the conceptual framework of the action model/change model schema discussed in Figure 3.5 of Chapter 3 to stakeholders. Then the evaluation team applied the forward-reasoning strategy, also discussed in Chapter 3, to facilitate stakeholders in articulating their action model and change model as follows.

Action Model

As discussed in Chapter 3, the action model consists of six components: implementing organizations, implementers, community partners and associate organizations, ecological context, intervention and service delivery protocols, and target population.

During the discussions, stakeholders raised a question: Should all of the many proposed changes be conceptualized as interventions? If not, how could

Figure 13.1 Logic Model of the Diversity Enhancement Program

stakeholders identify which change was an intervention? For example, should cultural change be identified as an intervention? To answer the question, evaluators introduced the distinct concepts of an intervention and a system change. *Intervention* refers to direct services provided to clients to change their knowledge, skills, and/or behaviors (desirable outcomes). *System change* is a modification of the contextual factors of a system in order to sustain an intervention and/or make an intervention more effectively affect outcomes. According to these definitions, cultural change is not an intervention. Rather, it is a system change relating to the component of ecological context. However, cultural change served important functions, such as providing disadvantaged/diverse students with a comfortable and supportive learning environment, that might indirectly influence their performance. With this distinction in place, stakeholders were able to identify the following components as the interventions of their project: develop a diversity-enhanced curriculum and provide mentors who could serve as role models, counseling and academic support, and scholarships/stipends to diverse and disadvantaged students. The action model's components and elements were as follows:

Implementing organization. The implementing organization was the project team that would initiate and coordinate activities.

Implementers. The project team would engage other faculty at the college in supporting the project and participating in its implementation.

Associate organizations/partners. The project team would seek support from the college's administration. Community organizations would collaborate in the recruitment of diverse and disadvantaged students; inter-professional partners would be invited to speak at the college and to help address health equity and social determinant issues.

Ecological context. The project would enhance the college's capacity to address health equity and social determinant issues and effect cultural change on campus.

Intervention and service delivery protocols. These protocols comprised a diversity-enhanced curriculum; mentors; and the provision of counseling, academic support, and scholarships/stipends to diverse and disadvantaged students.

Target population. The target population was diverse/disadvantaged students who were or could potentially be enrolled in the nursing program at the college.

A major purpose of the action model was expressed by the target population component, that is, to increase the enrollment of diverse and disadvantaged students. Figure 13.2 illustrates how other components were incorporated to contribute to reaching this goal. As indicated in the figure, the project team in the left-top box was the driving force for organizing and coordinating the project. They were responsible for recruiting other colleagues and to participate in the project, as indicated in the left-bottom box. As expressed in the middle-top box, the project team and faculty would seek support from the college's leadership, invite inter-professional partners to the college to speak and discuss health equity and social determinants, and collaborate with community-based organizations in recruiting diverse/disadvantaged students. The above efforts would create cultural change at the college and enhance the institution's capacity to address health equity and social determinant issues; this is indicated in the middle bottom box. These efforts would facilitate the development of intervention elements, including a diversity-enhanced curriculum; mentorship, counseling, and academic support; and scholarships, as illustrated in the right-top box. With these changes, the project would increase enrollment of diverse/disadvantaged students; see the right-bottom box.

Stakeholders said that the action model made explicit their assumptions and articulated their ideas about the intervention and system change components of the project very well.

Change Model

After completing the action model, stakeholders and evaluators moved on to discuss the change model. Stakeholders understood that the action model provided the necessary energy to drive and environment to support the change model. As discussed in Chapter 3, a change model consists of a minimum of three components: intervention, determinants, and outcomes. The intervention was already clarified by the action model. Stakeholders agreed that the project had three major outcomes: increase the retention, graduation, and employment of diverse and disadvantaged students. They also recognized that these three outcomes could be refined so as to classify retention as an intermediate outcome and graduation and employment as ultimate outcomes. Stakeholders invested more time in discussing what determinants were expected to mediate between the interventions and outcomes. The following two determinants were identified: (a) increasing diverse/disadvantaged students' self-efficacy and clinical skills and (b) increasing their grades and other performance indicators. The change model of the project is illustrated in Figure 13.3.

The entire action model/change model schema is illustrated in Figure 13.4.

Stakeholders expressed satisfaction with the action model/change model schema and commented that it accurately reflected the project's intention to remove contextual and environmental barriers that discourage or hinder diverse/disadvantaged students from joining the nursing workforce. The project had so many components and elements that stakeholders had trouble describing relationships among them in a meaningful way. The action model/change model schema helped them to synthesize the components and

Figure 13.2 Action Model of the Diversity Enhancement Project

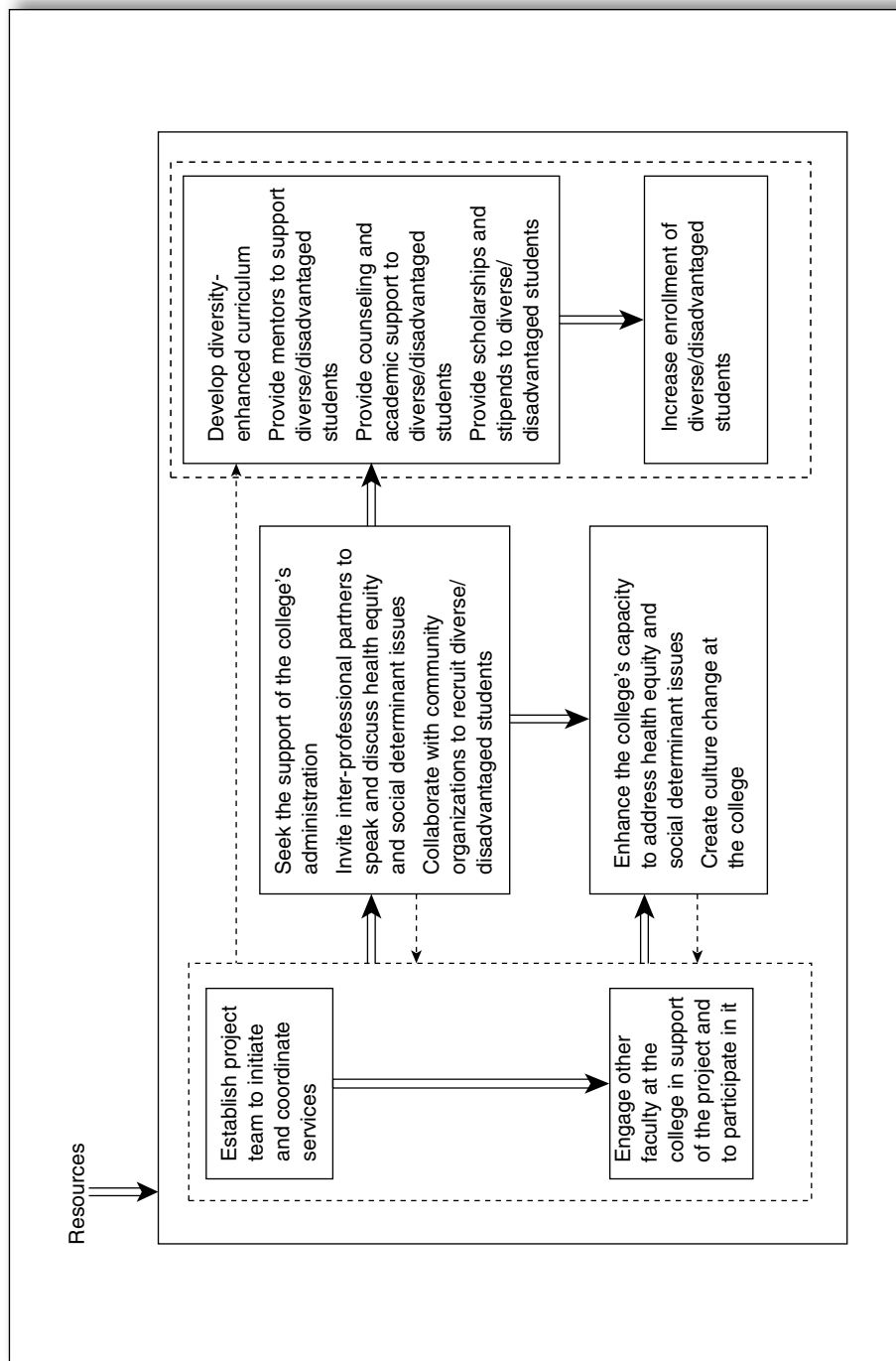
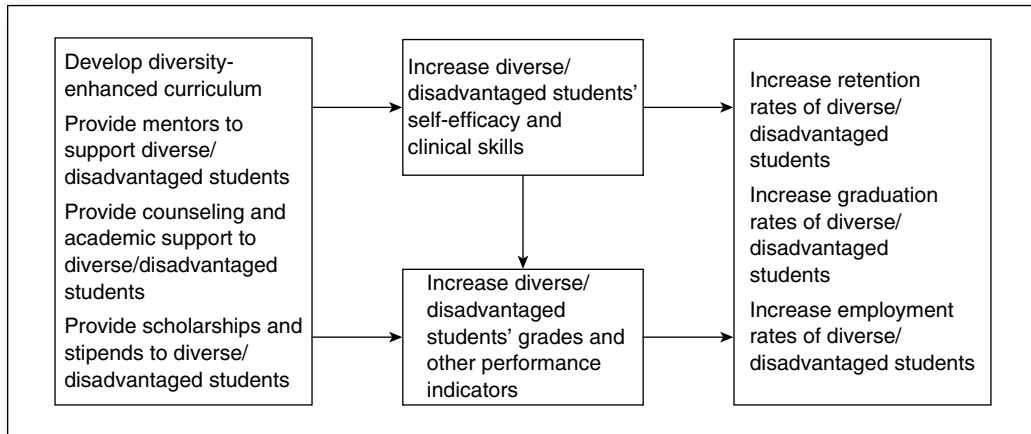


Figure 13.3 The Change Model of the Diversity Enhancement Project

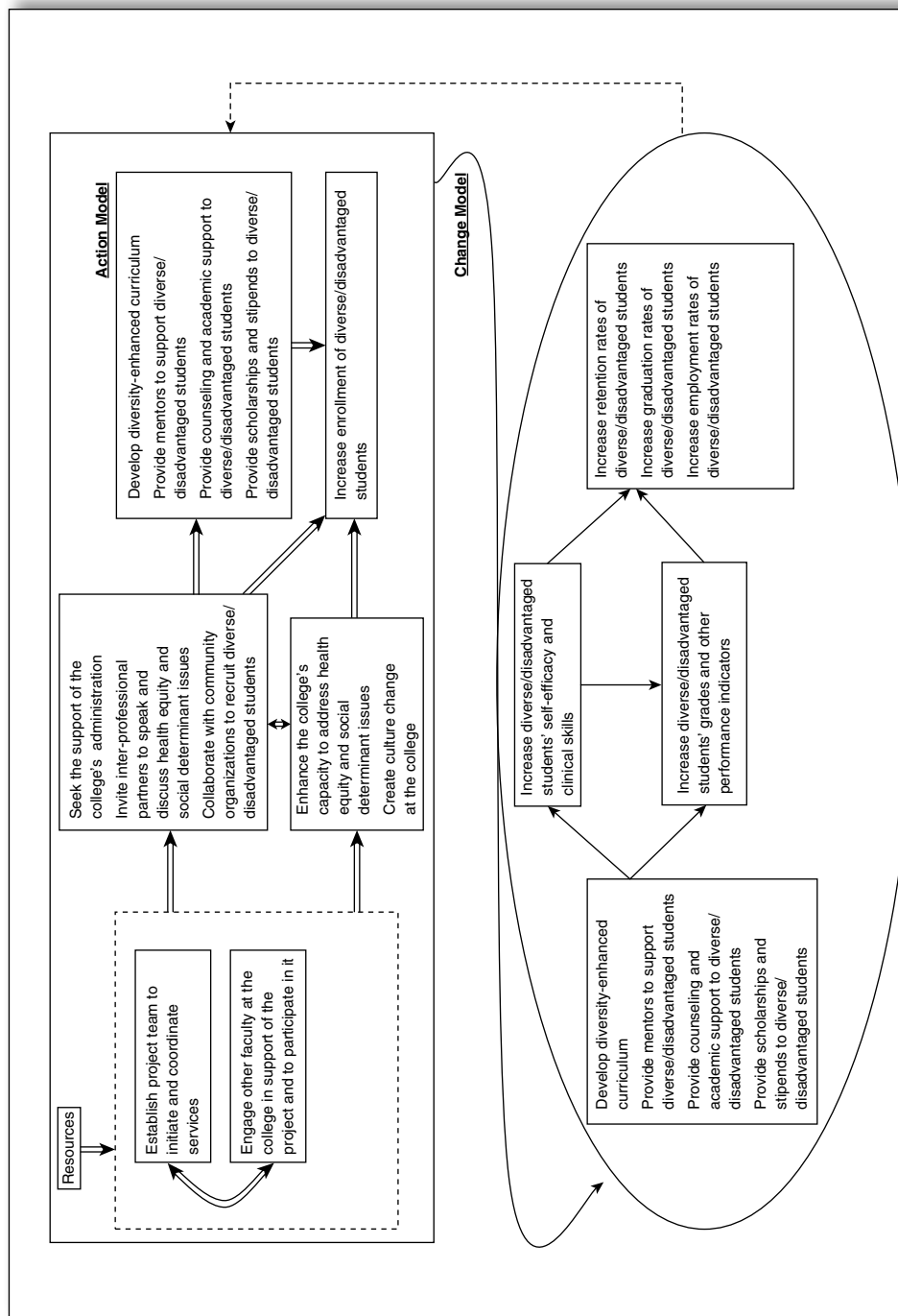
elements and describe their relationships in a systematic and coherent structure. Stakeholders also said they were particularly impressed that the schema made a distinction between the action model and change model. This distinction gave them insight into how to conceptualize the intervention and provided them with more insights into the different functions served by the components of the project and how they related to each other. The schema gave them many ideas about how to further strengthen the entire project. They felt that both the logic model and the action model/change model schema were useful: The logic model provided a quick summary of their project, while the action model/change model schema described their project in more depth and provided more insight to guide their planning and evaluation of the project. They planned to use both the logic model and the action model/change model schema for internal and external communication, and they believed they might use both when applying for funding in the future.

A COMMUNITY HEALTH INITIATIVE

Description of the Project

A community-based organization in a northeastern suburb of a US city launched a community health initiative to promote active living and healthy eating among community residents. A core feature of the initiative was to

Figure 13.4 The Action Model/Change Model Schema of the Diversity Enhancement Project



facilitate partnerships as a foundation for planning and implementing the initiative. Community partners were invited to collaborate in adapting and implementing evidence-based strategies to improve community health and prevent obesity. Key stakeholders believed that community partnerships were essential to implementing sustainable community change. The initiative began with the following activities:

- Partners from the community were invited to brainstorm and plan program activities.
- An awareness campaign was conducted that included online announcements, brochures, and media releases about the initiative.
- A needs assessment was conducted to understand community residents' eating and exercise habits and their views on health-related issues.

The key stakeholders expected that these activities would create momentum that would eventually lead to the overarching goal of increasing residents' healthy eating and active living.

Applying the Logic Model

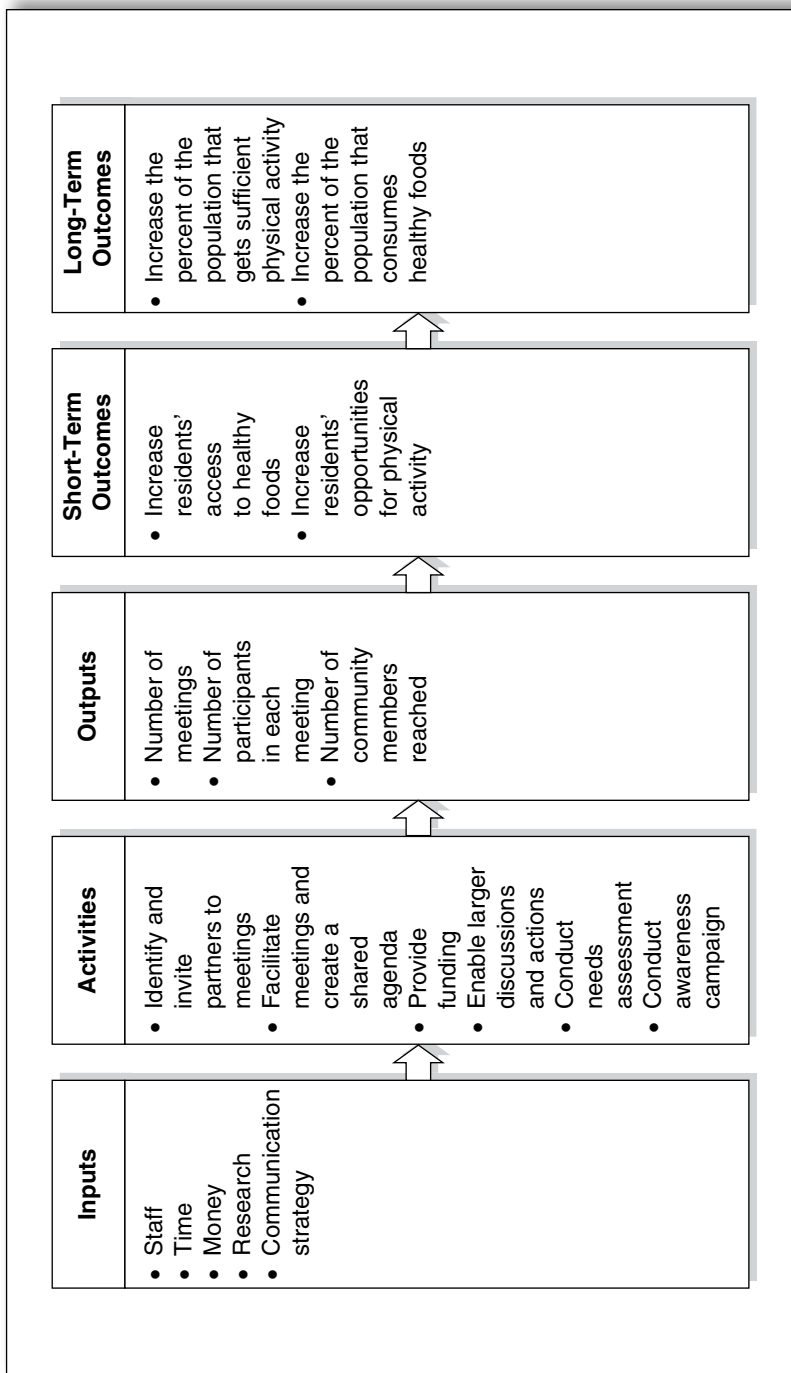
Before the launch of the initiative, key stakeholders and consultants had developed a logic model to describe the effort's major components. The logic model was developed by closely following the framework found in the literature (Wyatt Knowlton & Phillips, 2013; United Way of America, 1996). One year later, an evaluation team, which was not composed of the original consultants, was invited to plan and conduct an outcome evaluation of the initiative. The evaluators reviewed relevant program information, including the logic model, as they worked to plan an ongoing evaluation. Along with key stakeholders, the evaluators agreed to use the logic model previously developed as a basis for communicating about program planning and evaluation issues. The logic model is illustrated in Figure 13.5.

The partnership was characterized by activities such as identifying and inviting partners to meetings, facilitating meetings and creating a shared agenda, and providing funding to enable larger discussions. According to stakeholders, the partnership was a driving force for affecting subsequent components.

Applying the Logic Model to Planning Evaluation

Evaluators and stakeholders all agreed that the logic model was useful for identifying the major components of the initiative and organizing them in a systematic way that greatly facilitated communications about the effort's

Figure 13.5 The Logic Model of the Community Health Initiative



general themes. However, when the discussions began to go into more specific areas of program planning and evaluation, several issues arose, revealing that a typical application of the logic model was insufficient for serving the needs of the evaluation. These issues included the following:

1. *Should the partnerships promoted by the project be conceptualized as an intervention in the outcome evaluation?* When designing an outcome evaluation, it is essential to first identify the intervention(s) of a program (i.e., the outcome of what?). While designing an evaluation to assess this initiative's effects on health outcomes, key stakeholders and evaluators reviewed the logic model together to discuss which element(s) to consider the intervention. Key stakeholders said that partnerships could be regarded as an intervention for health outcomes for the purposes of evaluation for two reasons: (1) partnerships were a foundation of this initiative, and (2) the arrows connecting the input components to outcome components in the logic model clearly indicated that partnerships would affect outcomes. The evaluators agreed with stakeholders that partnerships were central to the initiative because they would allow for collaboration among organizations that could then facilitate changes in community health outcomes. However, the evaluators hesitated to agree that partnerships should be regarded as an intervention for purposes of evaluation. To qualify as an intervention component of a health promotion or social betterment program, the element must provide services to clients or must offer incentives or sanctions in some way to promote changes in behavior. While partnerships relate to and may facilitate interventions (e.g., a partnership can provide a basis for selecting or developing interventions that better serve a community), they do not by themselves provide direct services to any resident. Thus, partnerships might not meet the requirements to be considered an intervention in this community health initiative.

2. *Should the evaluation assess the initiative's population impacts?* According to the logic model, the long-term outcome of the initiative was to increase the proportion of the population that participated in sufficient physical activity and the proportion of the population that ate enough fruit and vegetables. However, the evaluators hesitated to rush into designing a population-based impact evaluation. Firstly, a population-based impact evaluation would be very expensive. Secondly, to meaningfully design and conduct a population-based impact evaluation, evaluators must first identify a population-based intervention. As discussed above, the logic model (Figure 13.5) did not provide sufficient information about the key intervention of the initiative. Without identifying the intervention, evaluators had no way to determine

whether the intervention would be population based or not. Literature on evaluability assessment (Wholey, 2004) suggests using logic models for evaluability assessment. However, after a few meetings, evaluators and stakeholders were still unclear about whether the population-based impact of this initiative was evaluable. The use of the logic model in this case was insufficient to clarify the issue.

3. *What were the roles of the awareness campaign and needs assessment?*

The initiative included an awareness campaign and a baseline needs assessment to understand community residents' behaviors and attitudes about healthy living. Stakeholders looked to the evaluators to help articulate the role of these two activities in the initiative. The logic model lumps these two items together in the *activities* component, thus providing limited information to help in understanding the separate functions of these items.

The evaluation team conjectured that the reason for these difficulties might be that the initial logic model developed by the consultants and stakeholders was a weak model. Efforts were made to develop a new version of a logic model. The new version, however, was no more effective than the initial version, illustrated in Figure 13.5. Problems still existed. Discussions proceeded without participants' reaching mutual understanding and agreement. However, additional questions were generated about what the intervention was and how to design an evaluation to meet stakeholders' needs.

As key stakeholders and evaluators struggled to use the logic model to guide the evaluation, the design of the evaluation study was delayed and stakeholders were stymied in planning future activities. Stakeholders and evaluators began to discuss the need to find a supplemental tool or alternative to the logic model to better articulate the dynamic processes and mechanisms underlying the initiative. A member of the evaluation team suggested using the action model/change model schema as an additional tool, and the stakeholders accepted this suggestion.

Applying the Action Model/Change Model Schema

Clarifying the Action Model/Change Model Schema

Evaluators explained the action model/change model schema in Figure 3.5 and the presumed relationship between the action model and change model, as discussed in Chapter 3. After stakeholders were familiar with the action model/change model schema and its components, the evaluators facilitated

stakeholders in clarifying the action model and change model underlying this initiative as follows:

1. *Clarify the relationship between partnership and intervention.* After evaluators introduced the action model and its components, key stakeholders were able to promptly identify their partnership effort as relevant to the action model. It related to the component of *associate organizations* and/or community partners. They were able to articulate that partnership is a platform for individuals and organizations to use to get together to discuss community needs and existing city and community-based organizations' services and to collaboratively plan and further advance community health outcomes, especially in the areas of healthy eating and active living. Stakeholders agreed that partnerships were an essential contextual factor in the initiative but, in itself, was not an intervention designed to directly change health outcomes. Rather, partnerships were an instrument by which partners could identify the interventions needed in the community, organize and coordinate these interventions, and implement them.

2. *Clarify the roles of the needs assessment and community awareness campaign in the initiative.* Using the logic model, stakeholders had difficulty articulating the roles of the needs assessment and community awareness campaign in the initiative. The action model helped them to articulate that the needs assessment provided partners with information about community needs and existing resources. This information was essential for partners to plan future actions that would address unmet needs. Similarly, stakeholders were able to express that the awareness campaign was relevant to the component of *ecological context* in the action model. The campaign's purpose was to increase community residents' awareness of health problems and to enhance residents' support for interventions that the initiative would propose in the future.

3. *Facilitate discussions to identify interventions.* Using the conceptual framework of the action model, stakeholders were able to clearly differentiate between partnerships and interventions. They were able to identify activities that provided direct services to residents, such as starting gardening clubs, forming walking groups, offering nutrition classes, and improving sidewalks, as the initiative's interventions.

4. *Clarify that intervention effects would be outcomes for individuals rather than impacts on the population.* Having identified the interventions, evaluators and stakeholders agreed that the effects of the interventions were more likely to materialize as individual outcomes—such as increasing physical activity and consumption of fruits and vegetables among participating

community members—rather than as impacts on the entire community, as had been stated in the initial logic model (see Figure 13.5).

Based upon the above clarifications, evaluators were able to facilitate stakeholders in developing their action model, illustrated in the top portion of Figure 13.6.

The cornerstones of the initiative are represented by three components of the action model on the left side of Figure 13.6:

1. *Build partnerships.* Stakeholders had envisioned the project as a community-wide effort. The identification and recruitment of community partners for participating in planning the initiative were essential. Activities related to this component have included: identifying and inviting participation from key partners, organizing representatives to collaborate on planning, scheduling meeting dates and locations, and planning meeting agendas.

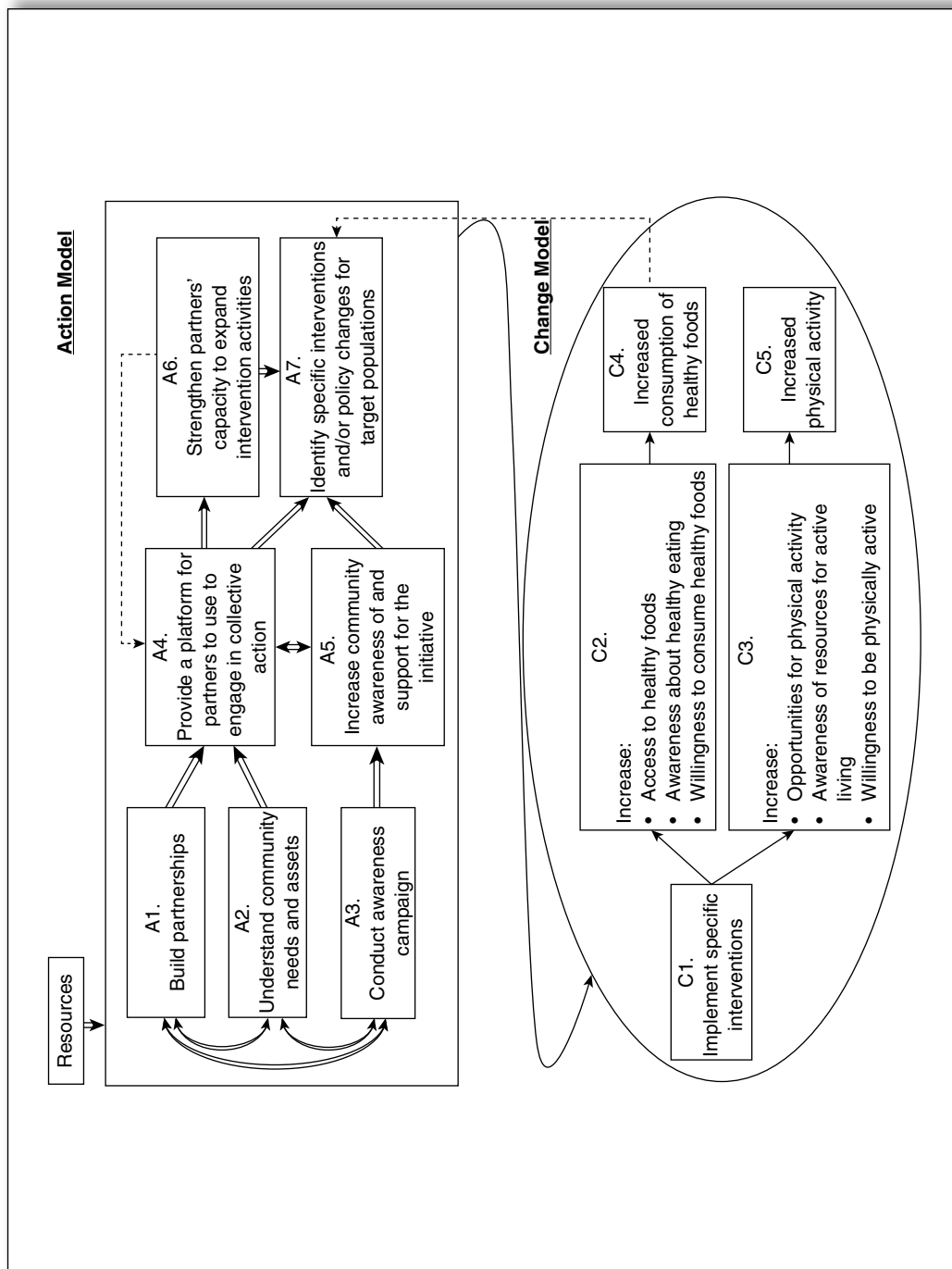
2. *Understand community needs and assets.* For partners to engage meaningfully in the planning of the initiative's activities, they had to first understand what the community's unmet needs are as well as what assets already existed that could be built upon to address these unmet needs. A needs assessment was conducted, in part, to serve this purpose. Information gathered from the needs assessment would be used to plan for the most appropriate and promising interventions to serve targeted populations.

3. *Conduct an awareness campaign.* Several marketing and communication efforts were conducted to increase community awareness of the initiative and its goals. These efforts included print advertisements and editorials in local newspapers and magazines, online marketing, signage, speaking engagements at local organizations, and sponsorships at health-related events in the community. This component relates to the ecological context of the initiative, discussed in the next section.

As illustrated in Figure 13.6, stakeholders articulated that these three components would be useful to achieve a number of shorter-term outcomes, including the following:

1. *Provide a platform for partners to engage in collective action.* Stakeholders believed that partnerships would serve as a platform that individuals and organizations could use to discuss, plan, and engage in collective action. That is, community partners would brainstorm ideas and strategies for community action through regular meetings, webinars, and online communication. Through participation in these meetings, partners would be provided the opportunity to

Figure 13.6 The Action Model/Change Model Schema of the Community Health Initiative



exchange ideas, share information and resources, and reinforce each other's commitment to the initiative and its goals.

2. *Increase community awareness of and support for the initiative.* Stakeholders believed that the awareness campaign would lead to increased resident awareness of the initiative. The supportive environment would encourage partners' brainstorming or planning activities.

3. *Strengthen partners' capacity to expand intervention activities.* Stakeholders believed that partners would benefit from increased knowledge and skills, the sharing of information and resources, and opportunities to collaborate. The experience would lead to increases in partners' capacity to plan new services or expand existing services provided by their organizations.

4. *Identify specific interventions and/or policy interventions for targeted populations.* Stakeholders wanted to work with partners to identify specific interventions they could launch to serve targeted populations, organizations that could coordinate the implementation of these interventions, and implementers who could deliver the services.

Furthermore, as indicated in Figure 13.6, stakeholders posited that a reciprocal relationship existed between the component of *platform for partners to use to engage in collective action* (A4) and the component of *partners' capacity to expand intervention activities* (A6). For example, when partners have strengthened their capacity for intervention activities, they may be more likely to participate in the partnership and work with others on collaborative projects.

Clarifying the Change Model

After clarifying the interventions, stakeholders were better able to articulate their change model. The components of the change model were as follows:

Specific interventions. Stakeholders envisioned that certain interventions would improve access to opportunities for physical activity and/or healthy eating. These might include starting community gardening clubs, forming walking groups, offering nutrition classes, enhancing parks, installing bike racks, and repairing walkways.

Determinants. Exercise-related interventions were intended to change the determinants by increasing opportunities for physical activity, increasing residents' willingness to exercise, and increasing residents' awareness of

resources for active living. Interventions related to healthy eating focused on determinants such as increasing awareness of healthy foods, increasing access to healthy foods, and so on.

Outcomes. The outcomes of these interventions included increased consumption of healthy foods and increased physical activity.

As articulated, the components of the action model would influence the performance of the change model. Partnership and needs assessment would better inform partners about interventions demanded by the community and increase their opportunities to collaborate on developing and delivering these needed services. Similarly, the success of the awareness campaign should motivate residents to participate in interventions. Furthermore, information from the change model would benefit the action model. For example, the evaluation of the effectiveness of these interventions would provide useful information that partners could use to plan future collaborative actions. The diagram of the change model is illustrated in the bottom of Figure 3.6.

The evaluation team and key stakeholders held three working-group meetings to complete a draft of this action model/change model schema. The draft was then presented via a webinar and to a larger group at an in-person meeting to refine the draft. Community partners actively participated in the development process. For example, in one of the meetings, stakeholders proposed to label each component in the action model with a letter A and a number to facilitate discussion. Similarly, the components in the change model are labeled with the letter C and distinct numbers. The suggestion was adopted as shown in Figure 13.6.

Contributions of the Action Model/Change Model Schema to the Community Health Initiative

The community health initiative described here is an ongoing project. In the current phase, the application of the action model/change model schema has allowed for two accomplishments:

1. *The action model/change model schema provides stakeholders with a coherent and thorough description of the initiative.* The action model/change model schema helps stakeholders to articulate and share a coherent view of the initiative, allowing for clearer elaboration of the roles of partnerships, needs assessment, and the awareness campaign and for explanation of how these components relate to other components, including interventions. Stakeholders' perspective on the action model/change model schema development experience was illustrated by a key stakeholder's comment: "Now I have a much better

idea about what we are doing and where we are going.” Stakeholders have taken ownership of the action model/change model schema and have used it for internal and external communications on project-related matters.

2. *Stakeholders and evaluators have been using the action model/change model schema to discuss the focus of the evaluation.* Given the action model/change model schema, stakeholders requested that the evaluation team conduct an evaluation of the action model. Specifically, the action model indicates that partners’ participation in partnership activities (A4) would increase their capacity to expand intervention capacity (A6). Stakeholders wanted to assess the relationship between A4 and A6. They hoped to use the information to plan future partner activities. In addition, stakeholders identified interventions for outcome evaluation as specified in the change model.

A GUIDE TO PRODUCTIVELY APPLYING THE LOGIC MODEL AND THE ACTION MODEL/CHANGE MODEL SCHEMA

As discussed in Chapter 3, literature reports that logic models are useful for many programs. However, to be competent in the use of logic models, evaluators need to know not only when logic models can be applied but also when this common practice may encounter problems and how to address problems when they occur. This chapter provides useful information in these areas. Logic models do not work all the time. Studies of the diversity enhancement project and the community health initiative indicate that intervention programs that emphasize system change may not lend themselves to the typical application of logic models. If programs seek to implement an intervention and at the same time implement a system change, stakeholders and evaluators using logic models alone are likely to encounter problems or confusion when trying to adequately describe a program. As demonstrated in this chapter, these problems can be addressed by applying the action model/change model schema as a way to further elaborate the understanding of the program provided by the logic model. The schema uses different lenses to examine the initiative and therefore allows for clearer articulation of key stakeholders’ views. As a result, a more complete program description can be developed. The schema can also provide a better guide for program planning and evaluation design.

The general principle derived from the above discussion is that the action model/change model schema or other program theories can contribute to better logic models by clarifying a program’s underlying assumptions regarding contextual factors and causal mechanisms. Evaluators thus have at least two options to choose from in applying a logic model:

Option 1: A reactive approach for applying logic models and the action model/change model schema. With this approach, evaluators continue a typical application of logic models until problems are encountered. They then apply the action model/change model schema as an additional tool for addressing the problems that have arisen. In other words, evaluators may want to continue to apply a logic model as usual. If the model works well, there is no need to make any change. However, if the application of a logic model is not as useful as it is expected to be, then evaluators can apply the action model/change model schema as a supplemental tool in an attempt to address these limitations. The diversity enhancement project and the community health initiative discussed in this chapter are examples of such an approach.

Option 2: A proactive approach in which the action model/change model schema is used before logic models. Although Option 1 is feasible, it may be a disorganized or perhaps chaotic way to apply a logic model and the action model/change model schema. Programs that are designed to address both system change and intervention issues may not lend themselves to the use of a logic model alone. Knowing this, evaluators and stakeholders can assess whether their program falls into this category. If so, they can plan ahead to apply an action model/change model schema to clarify system change issues before using a logic model. Alternatively, they can plan to apply the schema alone to describe the intervention program, emphasizing the system change, if there is no requirement for stakeholders to apply logic models. As demonstrated in this book, the action model/change model schema is a self-sufficient tool for guiding program planning and evaluation.

Because the action model/change model schema is more comprehensive than logic models, a concern is whether evaluators using the schema would have more difficulty explaining the schema to stakeholders than if they used logic models. Stakeholders' and evaluators' experiences with the diversity enhancement project and the community health initiative indicate otherwise. Stakeholders in these programs felt they had no difficulty understanding the action model/change model schema and that the schema articulated their view well. They were able to promptly relate their program activities to the components of action model and change model schema.

SYSTEM CHANGE AND EVALUATION IN THE FUTURE

Community-based organizations and government agencies are highly concerned about the sustainability of their programs. We see this interest in, for

example, the community health initiative's emphasis on partnership and the diversity enhancement project's focus on removing environmental barriers. Government agencies likewise share a desire to see their funded programs thrive. Due to budget constraints, interorganizational collaboration has become a popular strategy by which organizations and agencies can do more with less. It is conceivable that more interventions will stress contextual factors or system change, such as interorganizational collaborations, or other components of the action model as an essential ingredient in an intervention program. In this regard, more studies are needed that study the conditions that favor or disfavor the use of logic models and examine how to extend the typical use of logic models by means of additional approaches, such as the action model/change model schema.

QUESTIONS FOR REFLECTION

1. In the diversity enhancement project, stakeholders believed that the logic model served which of their needs well? Which of their needs did it serve less well? Do you agree with their views? Why or why not?
2. What are the differences between an intervention and a system change? In the diversity enhancement program, which elements were the interventions? Which were system changes?
3. Provide two reasons why stakeholders are interested in system change.
4. Give your own examples of interventions and system changes.
5. Compare and contrast the logic model and the action model/change model schema in terms of how they address issues related to interventions and system changes.
6. Explain why stakeholders in the community health initiative initially viewed partnership as an intervention for health outcomes.
7. Identify three areas in which evaluators and stakeholders of the community health initiative felt the logic model was not as useful as expected in helping them clarify the issues with the program.
8. Do you think the evaluators of the community health initiative had an obligation to raise concerns about the stakeholders' ideas regarding assessment of the effects of the partnership on health outcomes? Give reasons supporting both sides of this question.
9. Why was the action model/change model schema able to clarify the confusion that existed in the community health initiative?

10. Do you prefer to conceptualize the action model/change model schema as an extension of or as an alternative to a logic model? Why?
11. Based upon these two cases, what kinds of intervention programs do you believe may favor or not favor the use of logic models? Why do you believe so?
12. Compare and contrast the reactive approach and proactive approach to applying a logic model and/or the action model/change model schema.
13. Identify the relative strengths and limitations of a logic model and the action model/change model schema.
14. According to the author, why might more intervention programs emphasize both intervention and system change in the future? Do you agree or disagree with this view? Why?

CHAPTER 14



Formal Theories Versus Stakeholder Theories in Interventions

Relative Strengths and Limitations

As discussed in Chapters 3 and 12, an intervention program could be developed on a foundation of either formal theory or stakeholder theory. To advance theory-driven evaluation, it is important for evaluators to understand the characteristics of these two kinds of program theory and their relationship to intervention programs and evaluation. This chapter will provide a thorough analysis of the issues related to these two kinds of program theory.

FORMAL THEORY VERSUS STAKEHOLDER-IMPLICIT THEORY AS A BASIS FOR INTERVENTION PROGRAMS

In developing an intervention program, the adopted program theory—and particularly the change model—can be based on either well-defined formal or stakeholder-implicit theory. The characteristics and merits of the two options require some discussion.

Intervention Programs Based on Formal Theory

When the development of a program is based on formal theory, it is a *formal theory-based program*—and it may be bursting with fertile information about which determinants

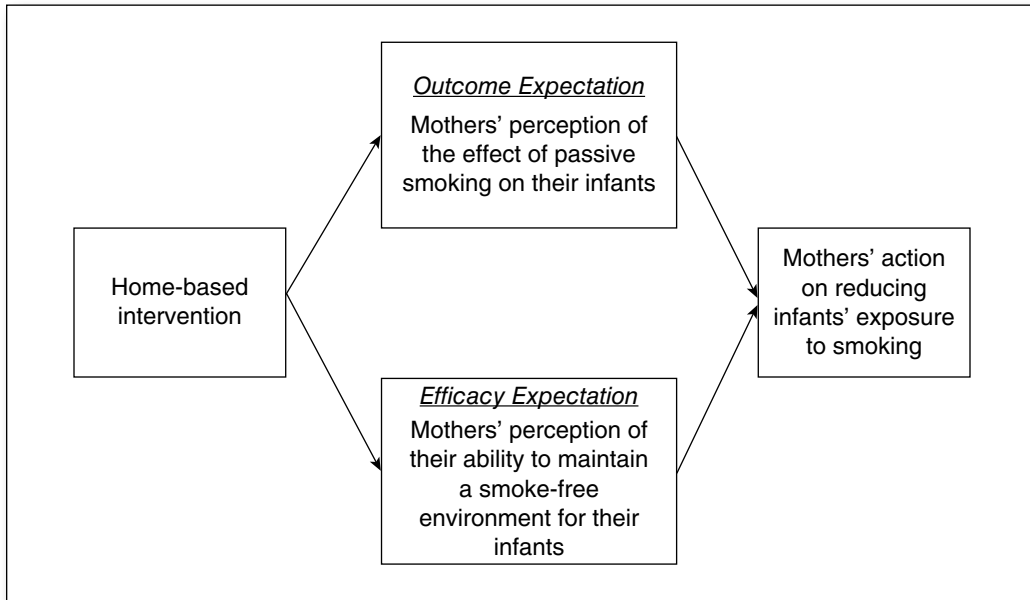
will make the program work. This is especially so of formal theory-based programs in the areas such as health promotion, education, criminal justice, and human services. The science available in these areas helps program designers and evaluators understand why a particular determinant should have the power to shape outcomes (Bartholomew et al., 2001; Witte, Meyer, & Martell, 2001). Formal theories are usually well tested, so choosing to base program designs upon their principles can eliminate trial-and-error searching for determinants. In fact, formal theory-based programs are often developed with scientists and scholarly researchers at the helm.

For example, with the goal of helping mothers to reduce passive smoking by infants, Strecher et al. (1989) applied Bandura's (1977) social learning theory to identify determinants and design an intervention. According to social learning theory, behavioral change (and the maintenance of it) arises from new expectations concerning outcomes of a behavior. This is called *outcome expectation*. Social learning theory is also concerned with *efficacy expectation*, which comprises beliefs concerning one's own capability to engage in the particular behavior. The home-based program was targeted at mothers and was designed with their outcome expectations *and* efficacy expectations in mind. In applying social learning theory to infants' passive smoking, the outcome expectation became mothers' perceptions of what happens to an infant exposed to environmental tobacco smoke. The efficacy expectation became mothers' perceptions about their own ability to create and maintain a smoke-free environment for their infants. Strecher and colleagues included outcome and efficacy expectations as two determinants of the intervention. Figure 14.1 illustrates the research team's formal theory-based program.

The theory-driven evaluation conceptual framework discussed in Chapter 12 is of value to those needing to assess formal theory-based programs. There are clear advantages to basing a program design on formal theory. The program will garner respect more readily. In addition, because scientists often take the lead in such programs, ideal testing conditions may be available for evaluations; this increases the chance of finding the thing that will make a program tick. But danger exists with formal theory-based programs as well. They tend to focus heavily on academic interests rather than stakeholder interests. Stakeholders often find that programs tied to formal theory are too controlled and fail to sufficiently reflect the real world.

Programs Based on Stakeholder Theory

A program founded mainly on stakeholder-implicit theory is a *stakeholder theory-based program*. The majority of intervention programs that operate in

Figure 14.1 A Formal Theory-Based Program to Reduce Infants' Passive Smoking

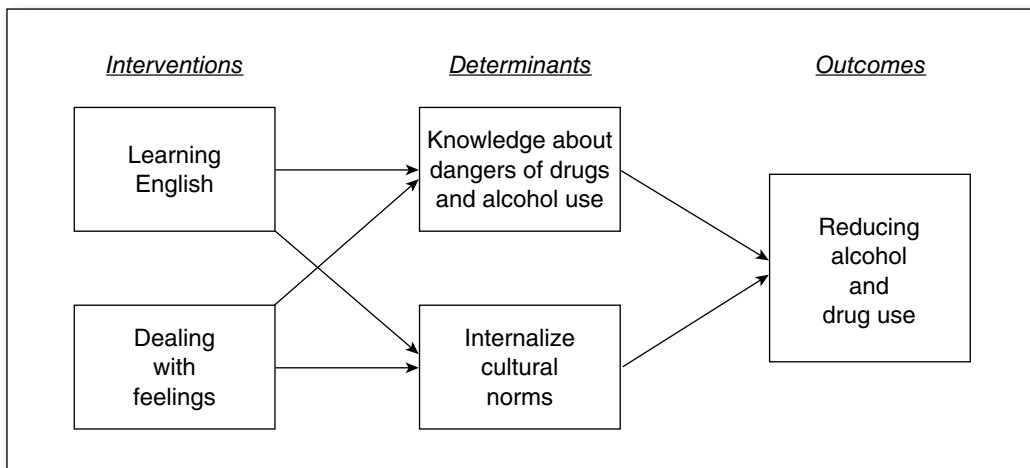
SOURCE: Adapted from Strecher et al. (1989).

a community have not been designed by scientists but rather by stakeholders such as program designers, program directors, and program staff. Each of these individuals perceives the nature of the problem in an individual way and develops a personal preference for a particular solution to the problems confronted. These perceptions and preferences may come from past experiences, conventional wisdom, discussion with peers, advice from experts, formal theories, acquaintance with similar programs—even hunches. In other words, stakeholders have their own program theory. Of course, stakeholder theory is not usually rendered as an explicit and systematic statement in the same way as formal theory. Stakeholder theory is implicit theory. It is not endowed with prestige and attention as is formal theory; it is, however, very important from a practical standpoint because stakeholders draw on it when contemplating their program's organization, intervention procedures, and client-targeting strategies. Stakeholders' implicit theories are not likely to be systematically and explicitly articulated, and so it is up to evaluators to help stakeholders elaborate their ideas.

An illustrative case is Clapp and Early's (1999) study of alcohol and drug abuse prevention programs for young Hispanic students. These programs tended to be vague as to their components and rationales. Clapp and Early met with focus groups of clients, staff, and parents from each program in order to clarify the stakeholders' implicit theories. In one school-based program, the researchers ironed out two intervention elements desired by stakeholders: English language acquisition and a course in coping with feelings. The stakeholder theory assumed that these two elements would sequentially affect the two program determinants: identifying with the larger culture and internalizing that culture's norms. When fully exploited, the stakeholders theorized, the two determinants would produce an outcome of reduced alcohol and drug abuse. The change model implicit in the stakeholders' beliefs is delineated in Figure 14.2.

The conceptual framework of program theory can help evaluators to "read" stakeholders' implicit theories in three ways. The first involves making the implicit explicit as well as systematic. If asked to participate from the very beginning of the program, an evaluator working from the conceptual framework sees places where the program would benefit from spelling out the stakeholders'

Figure 14.2 A Stakeholder Theory-Based Alcohol and Drug Abuse Prevention Program for Hispanic Students



SOURCE: Adapted from Clapp and Early (1999).

assumptions. The conceptual framework also suggests which assumptions might be weak or which elements of a theory are missing (see Chapters 4 and 5 for further discussion). The second way the framework helps is by saving time in situations in which feedback is needed quickly to confirm that a program design is on track. With timely feedback, a program or implementation can be developed to a level at which it will operate smoothly and efficiently. The conceptual framework offers a direction for evaluation design that, although ready-made, is also amenable to tailoring, a point elaborated in Chapter 6 of this book. Third and last, the conceptual framework's versatility allows evaluators to assess comprehensively the overall quality of implementation and program effectiveness, while at the same time identifying strengths and weaknesses in its elements (see Chapters 7 and 12).

Views on the Relative Value of Formal Theory-Based Interventions and Stakeholder Theory-Based Interventions

As discussed above, formal theory-based interventions follow the well-traveled path of social/behavioral science theories developed in academia, such as social learning theory or reasoned action theory. The use of this deductive approach for formulating program theory is familiar to researchers and evaluators (Christie, 2003). By contrast, stakeholder intervention theories mainly originate from stakeholders' ideas, observations, and experiences working with clients and partners in a community. Stakeholder intervention theories are implicit, inductive, and less systematic and coherent than are formal theories.

In terms of reputation and the priority they are given for funding and dissemination, these two interventions have fared differently. Researchers, academic institutions, and funding agencies often regard formal theory-based interventions as prestigious in terms of theoretical origin and potential effectiveness. Therefore, formal theory-based interventions are likely to receive funds for research or evaluation. In fields such as public health, researchers often apply randomized controlled trials (RCTs) or other traditionally rigorous methods to assess the efficacy of formal theory-based interventions. Because of evidence of efficacy, these formal theory-based interventions are regarded by researchers and many funding agencies as a top priority for dissemination. Also, results of evaluations of these interventions are likely to be published in journals.

Compared with their prestigious formal-theory family members, stakeholder theory-based interventions are the poor relation. These interventions are usually

not systematically studied. Researchers regard them as having low prestige in terms of theoretical origin and potential effectiveness. If these interventions are evaluated, the evaluation designs often use quasi-experimental or pre-experimental methods rather than experimental methods. Due to a lack of strong evidence, as evidence is traditionally appraised, to demonstrate their efficacy, stakeholder theory-based interventions are less likely to receive funds for study and to have follow-up discussions in the literature. Also, researchers and funding agencies tend to not recommend these interventions for dissemination.

In terms of real-world applications, a different picture emerges. Practitioners regard stakeholder theory-based interventions as practical and as having local relevancy. In spite of lacking rigorous evidence of stakeholder theory-based interventions' effectiveness, stakeholders favor such interventions and implement them in communities (Chen, 2010; Chen & Garbe, 2011). This is not the case with many formal theory-based interventions or evidence-based interventions, which are not often used in real-world practice. This chasm between science and practice has raised legitimate concerns across many fields, as discussed in the literature (Glasgow, Lichtenstein, & Marcus, 2003; Green & Glasgow, 2006; Wandersman et al., 2008). Researchers and funding agencies have proposed ways to narrow this gap. One popular strategy is to persuade communities to more frequently use formal theory-based interventions, as exemplified by the current evidence-based intervention movement (Donaldson et al., 2009). Stakeholders are offered incentives such as funds, technical assistance, or capacity building to adopt formal theory-based interventions. The assumption here is that formal theory-based interventions with evidence of efficacy are superior to stakeholder theory-based interventions in addressing community problems.

However, this assumption is questionable on several fronts. Take, for example, the issue of evidence in evidence-based interventions. Such so-called evidence is mainly some demonstration of efficacy produced in ideal and controlled settings. The majority of evidence-based interventions lack real-world evidence. We simply do not know how these interventions will work when ordinary community-based organizations attempt to organize, manage, and implement them, and we do not know whether such interventions can satisfactorily address real clients' problems in a real-world setting (Chen, 2010; Chen & Garbe, 2011). In fact, few evaluations or studies have empirically compared the relative merits of formal theory-based and stakeholder theory-based interventions in a real-world setting. Too many unanswered questions remain, including the following:

- Are formal theory-based interventions superior to stakeholder theory-based interventions in the real world?

- If so, what are the areas of superiority?
- If formal theory-based interventions are superior, why do stakeholders not enthusiastically embrace them?
- If stakeholder theory-based interventions are inferior, why do stakeholders frequently apply them in practice?

Despite the lack of evidentiary support, the assumption of the superiority of formal theory-based interventions has been frequently used as a basis to fund and promote them. If, however, this assumption of superiority is unfounded, then the preference given to such interventions could result in wasted money or even counterproductive efforts. The stakes are sufficiently high that it is important to explore the issue.

FORMAL THEORY VERSUS STAKEHOLDER THEORY: A CASE STUDY

A recent study of an anti-secondhand smoking program (Chen & Turner, 2012) provides insight into the relative value of formal theory and stakeholder theory. The rest of this chapter will provide a detailed discussion of the purpose of this study, its findings related to formal and stakeholder theory, and the findings' implications for conducting theory-driven evaluation.

Program Theory Underlying the Anti-Secondhand Smoking Program

The target population of the program was residents living in a public housing complex owned and managed by the local housing authority in a midsize Georgia city. Evaluators applied theory-driven evaluation to evaluate the program. The program's change model contained two interventions calculated to address exposure to environmental tobacco smoke. The first intervention, based on a formal theory, became known as *community health advisor* (CHA; Cornell et al., 2009; Martin et al., 2011; Wallerstein & Bernstein, 1994). Members of the planning committee who were affiliated with universities as instructors or doctoral students proposed this evidence-based intervention. The second intervention was proposed by other committee members who had extensive experience in delivering social services in low-income neighborhoods but who were not affiliated with a university. This stakeholder theory-based intervention was known as the *education/nonsmoking signage intervention*.

Proponents of CHA proposed to identify and train indigenous lay health advisors to improve the health of persons in the community (Fleury, Keller, Perez, & Lee, 2009; Wallerstein & Bernstein, 1994). They postulated that within every community, people in formal and informal social networks exchange health information and influence health behavior decisions. CHAs are social network members who are familiar with a particular community. When CHAs receive social cognitive theory-based training (Bandura, 1977)—including didactic methods, role-playing, motivational interviewing techniques, and ongoing mastery testing of the intervention curriculum—they have the capacity to raise community awareness of health problems and enhance community members' self-efficacy for changing health risk behaviors. Ideally, to build relationships and to deliver services, CHAs have multiple contacts within the community.

During the planning meeting, some members of the committee raised concerns about the feasibility of applying the CHA model to residents in a low-income housing complex. Training—and especially service delivery—is labor-intensive and time-consuming. Therefore, some committee members proposed the education/nonsmoking signage intervention as an additional option. This intervention aimed to educate residents about the harmful effects of tobacco smoke on family members, especially children. It would further persuade them to display policy signage, such as a window sticker and refrigerator magnet, stating, "This is a nonsmoking home." Thus, to reduce tobacco smoke exposure, residents could declare a smoking restriction policy for their apartments. The idea for this intervention came from stakeholders' experience working with low-income families in housing complexes and from their observations of how residents interact with tobacco-smoking guests.

The low-income residents of public housing in the community were largely single, young African-American women with children. Members of the planning committee observed that these young women had difficulty challenging the smoking behaviors of those visiting their homes such as parents, relatives, friends, and significant others. This difficulty might hinder residents' capacity to prevent tobacco smoke exposure at home. Displaying nonsmoking signs in a home could alleviate the problem. The sign would make a formal announcement to visitors that the household had a smoking-restrictive policy. When visitors attempted to smoke, the host could use the sign as an excuse for engaging in conversation about the apartment's nonsmoking policy. If many nonsmoking signs were displayed in a neighborhood, their high visibility could establish a social norm that restricted tobacco smoke in the neighborhood. This in turn could further affect individual decisions, leading to a reduction of environmental tobacco smoke in apartments. If a family member smoked, the decal could

assist discussions about the need for that smoker to reduce others' exposure to tobacco smoke, perhaps by smoking outside or by not smoking at all. Although the education/nonsmoking signage intervention's effectiveness was unsupported by evidence, stakeholders believed, based on their experience, that residents would be more receptive to this intervention than to the CHA intervention.

To increase the chances that the outcome would be achieved, key stakeholders decided to apply both interventions. We discuss in the next section the procedure for putting these two interventions into practice.

Action Model

An action model specifies factors or a context that stakeholders believe is essential for starting or supporting an intervention as discussed in Chapter 3. An action model includes the following components: implementing organizations, implementers, intervention/service delivery protocols, target population, and partners. An action model is useful for assessing the implementation process, understanding the relationship between an intervention and its environment, and providing a context in which to interpret the intervention's outcomes. With evaluators' assistance, stakeholders clarified their action model for both interventions of their program, as discussed below.

Target Population

The target population of the program was 1,785 residents in two Columbus Housing Authority complexes: Booker T. Washington (BTW) and Baker Village (BV). According to the baseline survey, 43% of the residents reported at least one tobacco user at home, and 57% of residents said they were exposed to secondhand smoke at home.

Intervention Protocols

Intervention protocols included the following components:

The Education/Nonsmoking Signage Intervention Protocol. The protocol of the education/nonsmoking signage intervention included a smoke-free kit with a set of antismoking brochures for residents, nonsmoking signage items, and a service delivery manual for outreach workers. Policy signage items were created in multiple forms, including window decals, door-handle decals, refrigerator magnets, and tabletop tents. All had the statement "This Is a

Non-Smoking Home.” The brochures for residents described and documented the harmful effects of environmental tobacco smoke. The manual for outreach workers contained the following practical information: Approach residents in the home; engage in conversation with the primary resident; use antismoking brochures to explain the harmful effects of environmental tobacco smoke, emphasizing the effects on family members; discuss the purpose of the nonsmoking signage; encourage residents to display multiple nonsmoking signage items at home to declare a smoking-restriction policy; explain the implications of the signage for guests who were smokers; deal with resistance if some family members were smokers; and help family members by locating smoking-cessation services if those family members wanted to quit smoking.

The training for the intervention took 3 hours. Delivery of the education/nonsmoking signage intervention was estimated to take 20 minutes.

The CHA Protocol. The 3-day CHA training curriculum (a total of 20 hours) concentrated on tobacco smoke and health issues and leadership development. The protocol covered activities for developing specific skills needed to implement the intervention. The purposes were to ensure CHAs were familiar with how to communicate with clients about issues associated with secondhand smoke issues and how to encourage clients to take action. CHAs learned and practiced skills through various role-playing activities and received feedback from other CHAs. Trainees were paid a stipend as an incentive to complete all the modules of the curriculum. After completing training, CHAs were to take an estimated 45 minutes per client visit to administer the intervention. The training stressed that CHAs should conduct multiple visits to monitor residents’ efforts to reduce secondhand smoke and to encourage them to establish a smoke-free home.

Mode of Service Delivery

The planning committee’s idea was to deliver the two interventions in sequential order. Because they expected residents would be more receptive to the education/nonsmoking signage intervention than to CHA, the education/nonsmoking signage intervention was delivered first. Trained outreach workers went to the target housing authority apartment complex to solicit participants. If an adult resident of an apartment agreed to participate in the program, the outreach worker provided the education/nonsmoking signage intervention. At the end of the intervention, the outreach worker asked participants if they were willing to receive the additional CHA intervention in the near future. If the resident agreed to the CHA intervention, the outreach

worker made an appointment with the resident for a follow-up visit or multiple visits to complete the intervention.

Implementing Organization and Partners

The implementing organization was a local church with a long history of community service. The church was located close to the housing authority complexes and had a good working relationship with the housing authority itself. Moreover, the church was well connected with other community-based organizations in the area. Project partners included a cancer prevention coalition, a local college, a clinic, an urban development group, a girl-empowerment organization, and the local housing authority.

Implementers

The housing authority partners recommended unemployed BTW residents as candidates for outreach project workers. These particular residents were readily available to work for the project and could use their natural networks to reach other residents to deliver the interventions. Unemployed residents were also attracted to the positions because their participation as outreach workers would allow them to claim community service hours—a requirement for staying in their apartments. The positions were announced at the complex. Criteria used to screen candidates included educational background, employment history, smoking status, relationships with neighbors (especially as concerned offering advice), and willingness to complete the program training requirements. Seven tenants met these criteria and were recruited for the project. They agreed to participate in the training to deliver the interventions. They would be paid \$10 per hour.

Ecological Context

Ecological context includes those contextual factors relevant to an intervention program. The planning committee initiated several community activities related to the program. Those activities included a talent show and a mother/daughter double-Dutch jump rope competition as well as a seminar devoted to a discussion of smoking and secondhand smoke. The activities were open to BTW residents and to the public. These community activities were intended to enhance public awareness of the health problems posed by secondhand or environmental smoke as well as to create an atmosphere conducive to acceptance of the interventions. Another contextual factor was a citywide effort to

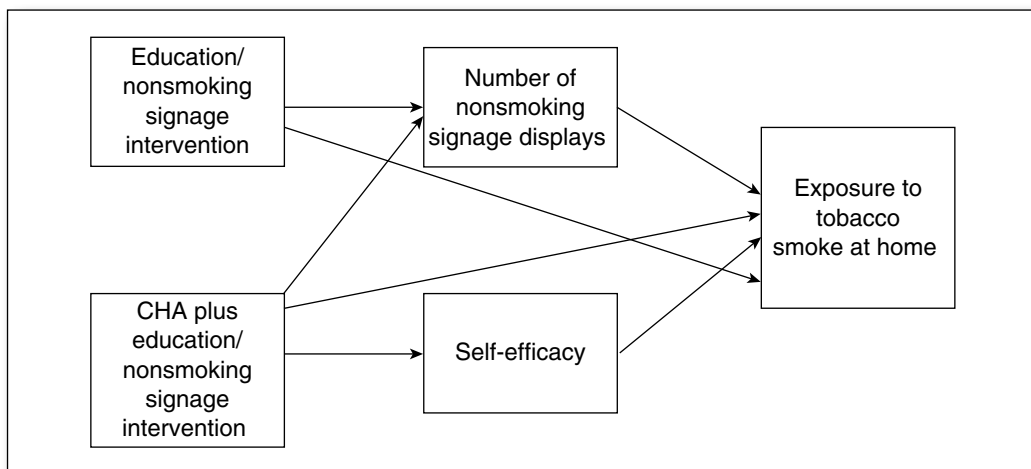
enact a clean indoor air ordinance, which might enhance BTW residents' willingness to participate in the project.

Outcome Evaluation Design and Change Model

A nonequivalent comparison group design (Shadish et al., 2002) was used to assess the effectiveness of the interventions. The research design measured the intervention variable for three groups: (1) the education/nonsmoking signage-only group, (2) the education/nonsmoking signage plus CHA group, and (3) the comparison group.

The change model of the program is illustrated in Figure 14.3. From left to right, the figure postulates two causal chains. The top portion illustrates that the education/nonsmoking signage intervention was expected to increase the number of nonsmoking signage displays, which, in turn, would decrease the extent of exposure to tobacco smoke. The bottom portion indicates that the enhanced intervention with CHA was expected to increase participants' self-efficacy in addressing the environmental smoke problem. That, in turn, would reduce the extent of exposure to tobacco smoke. Given that the enhanced intervention also had the education/nonsmoking signage intervention component, the intervention was expected to affect the number of nonsmoking signage displays. In addition

Figure 14.3 Causal Model of the Environmental Tobacco Smoke Prevention Program



to the indirect effects, both interventions were postulated to have a direct effect on tobacco smoke exposure.

The causal model would be useful to compare the causal chains of these two interventions. For example, if the education/nonsmoking signage intervention and CHA intervention worked equally well, both should have been able to activate their respective determinants. This, in turn, should have reduced tobacco smoke exposure. But if one intervention could activate the causal chain while the other failed to do so, then the former would be superior to the latter.

Process Evaluation Design

Process evaluation assesses how an intervention program is implemented. In this evaluation, stakeholders and evaluators were particularly interested in the implementers' and clients' experiences with or receptivity to the interventions. We collected the following data:

- *Training exit survey.* After training, outreach workers were asked to fill out an anonymous exit survey that asked about their satisfaction with the training and if they had suggestions for improvement.
- *Record of outreach and service delivery.* Outreach workers were asked to record the number of contacts they made, residents' acceptance or rejection of each intervention, and the length of each visit.
- *Outreach debriefing.* Supervisors documented weekly meetings with each outreach worker to discuss experiences with reaching clients and providing interventions. The supervisor used the debriefing to check outreach workers' service delivery against the protocols and to remind workers of the expectation of fidelity in implementation.

EVALUATION FINDINGS

Results of Process Evaluation

Process data indicate that the assumptions made in the action model held true during implementation except those concerning outreach worker training and client recruitment.

1. *Training outreach workers.* Seven trainees attended 3 hours of training for the education/policy signage intervention and 20 hours of training for the

CHA intervention. According to the exit survey, the training on the education/policy signage intervention went well. The trainees particularly liked the role-playing exercises in which they practiced recruiting residents to participate in the intervention. However, the training for the CHA intervention did not go as smoothly. Trainees were frustrated with the lengthy protocol and challenges of leadership skills. Two trainees quit the project due to frustration. The remaining five trainees stayed with the project and agreed to serve as outreach workers to deliver both interventions.

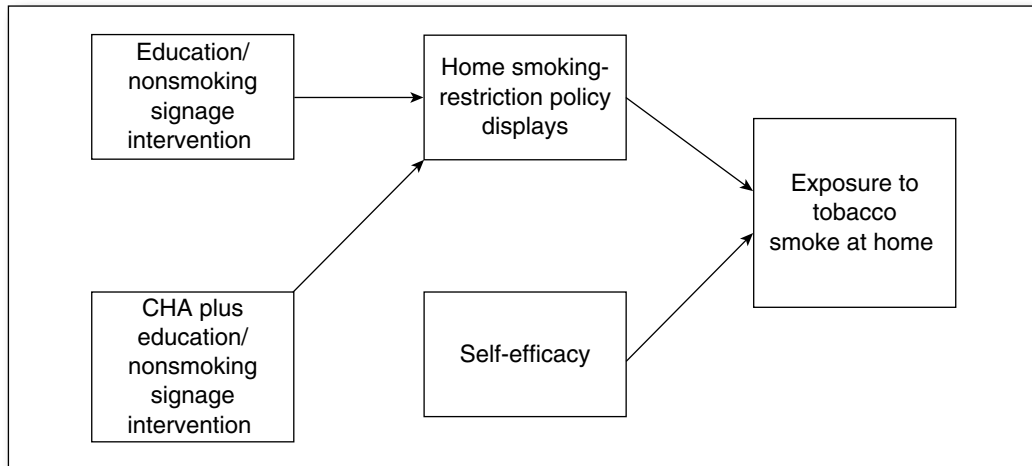
2. *Participants' receptivity to the interventions.* Among 138 participants in the intervention group, 104 (75%) participated in the education/nonsmoking signage intervention. The other 34 (25%) residents agreed to participate in the education/nonsmoking signage intervention and CHA. Furthermore, only 12 residents of these 34 (35.3%) agreed to more than one CHA visit. Debriefing data from five outreach workers indicated that participants were more receptive to the education/nonsmoking signage intervention than to the CHA.

Outcome Evaluation Results

It was found that, as predicted, the two intervention variables of education/nonsmoking signage intervention and CHA plus education/nonsmoking signage intervention significantly increased the number of nonsmoking signage displays at home. The CHA plus education/nonsmoking signage intervention was expected to affect self-efficacy, but this expectation was not supported by the data. As expected, two determinants (the number of nonsmoking signage displays and self-efficacy) significantly decreased the extent of exposure to tobacco smoke at home. In addition, the education/nonsmoking signage intervention variable and nonsmoking status variable also reduced exposure to tobacco smoke.

Using these findings, we revised the causal model of the interventions. Figure 14.4 indicates that the education/nonsmoking signage intervention translated effectively through the causal chains, as stakeholder theory had expected. The intervention affected the determinant (the number of nonsmoking signage displays), which, in turn, reduced exposure to tobacco smoke. On the other hand, Figure 14.4 indicates that while the CHA plus education/smoking signage intervention was theoretically expected to affect its determinant (self-efficacy), this linkage was not found. However, self-efficacy was found to affect the extent of exposure to tobacco smoke as predicted. There were no direct effects from the two intervention variables on the extent of exposure to tobacco smoke.

Figure 14.4 Revised Causal Model of the Environmental Tobacco Smoke Prevention Program, Based on Path Analysis



RELATIVE STRENGTHS AND LIMITATIONS OF FORMAL THEORY-BASED INTERVENTION AND STAKEHOLDER THEORY-BASED INTERVENTION

The findings from the case study provide insight into the relative strengths and limitations of formal theory-based and stakeholder theory-based interventions in terms of planning, implementation, and outcomes in the real world.

Theoretical Sophistication and Prior Evidence

This case study found that a formal theory-based intervention was superior to a stakeholder theory-based intervention in terms of theory sophistication and prior evidence supporting the theory. As demonstrated in this study, CHA is based on social cognition theory—a well-known formal theory. Both the theory and CHA are well studied and well respected in academic circles and are attractive to researchers and funding agencies.

On the other hand, the stakeholder theory-based intervention found in this study does not have these advantages. Stakeholder theory originates from stakeholders' ideas, experiences, and observations. These theories are hardly

studies and are rarely if ever discussed in the literature. They are often regarded as commonsense or informal theory. As demonstrated in this case study, neither the stakeholder theory nor the education/nonsmoking signage intervention had been formally studied. No prior evidence supported them. Because it lacks a theoretical foundation valued in academia and prior rigorous evidence, this intervention is not as attractive as CHA to researchers.

Efforts to Clarify the Change Model and Action Model in Program Theory

The literature explicitly discusses formal theory and formal theory-based interventions. Thus, evaluators can use the literature to construct a change model. On the other hand, a clear delineation of a stakeholder theory underlying a stakeholder theory-based intervention is usually not available in literature or documents. Evaluators must expend considerable effort to facilitate stakeholders in clarifying such theories. As shown in this case study, the literature discussed the theory underlying CHA intervention but did not formally discuss the education/nonsmoking signage intervention. In terms of an action model, the intervention protocol component tends to fare better in formal theory-based interventions, with a protocol being more likely to accompany a formal theory-based intervention. Evaluators can use the protocol to assess the implementation process. As shown in this study, CHA has a protocol for training and implementation. On the other hand, a stakeholder theory-based intervention may not have a well-developed protocol, and evaluators may need to facilitate stakeholders in developing one for evaluation purposes. For example, in this study, evaluators facilitated stakeholders in clarifying their education/nonsmoking signage intervention theory and developing an intervention protocol.

Efficacious Evidence Versus Real-World Effectiveness

As discussed in the introduction to this chapter, an intervention that proves efficacious in an ideal and controlled setting will not necessarily be effective in the real world (Chen, 2010; Chen & Garbe, 2011). Many evidence-based interventions are supported only by evidence of efficacy. Thus, whether such interventions are effective in the real world is not known. This case study provides some evidence to support the idea that efficacious interventions are not always effective. For example, as indicated in Figure 14.4, CHA was found to be ineffective

at activating its determinant (self-efficacy) in this real-world program. Its ineffectiveness may be attributed to reasons such as implementation problems (e.g., a lack of fidelity in service delivery or characteristics of the unemployed workers used as CHAs). Nevertheless, stakeholders view implementation problems as a weakness of using formal theory-based interventions or evidence-based interventions in a real-world context. They regard difficulties in hiring highly qualified, highly motivated staff or maintaining fidelity in the real world as a limitation of these interventions.

As discussed in the introduction, the lack of prior evidence substantiating efficacy or effectiveness limits stakeholder theory-based interventions. But the lack of evidence may reflect the low priority given to studying and evaluating such interventions rather than their ineffectiveness. Not all stakeholder theory-based interventions are ineffective. The findings of this evaluation provide some support for that argument by showing that the education/policy signage intervention was effective. More studies in stakeholder-based interventions are needed.

Viability

Evidence of efficacy is the major criterion for assessing the merits of an evidence-based intervention. This criterion may satisfy the desire for rigorous evidence, but it is too narrow to adequately reflect stakeholders' views. To better reflect stakeholders' interests, both viability and effectuality should be included in the scope of evaluation (Chen, 2010; Chen & Garbe, 2011). *Viability* refers to an intervention that is viable in the real world. More specifically, viability means stakeholders regard an intervention as practical, suitable, affordable, evaluable, and helpful. Because stakeholders are responsible for organizing and implementing an intervention, they are concerned about that intervention's viability. Researchers are less likely to be interested in viability issues. When researchers develop a formal theory-based intervention, their main interest tends to be more in theoretical sophistication and methodological rigor than in practicality or service deliverability. The evaluation of the case study provides some support for the argument that research-supported theory is not concerned with practicality, as it showed that two out of seven trainees were so frustrated with CHA training that they quit the project after the training.

When stakeholders propose an intervention, they tend to factor in components such as implementers' skills, the community organization's capacity, cultural competence, resource availability, and clients' preferences. Since

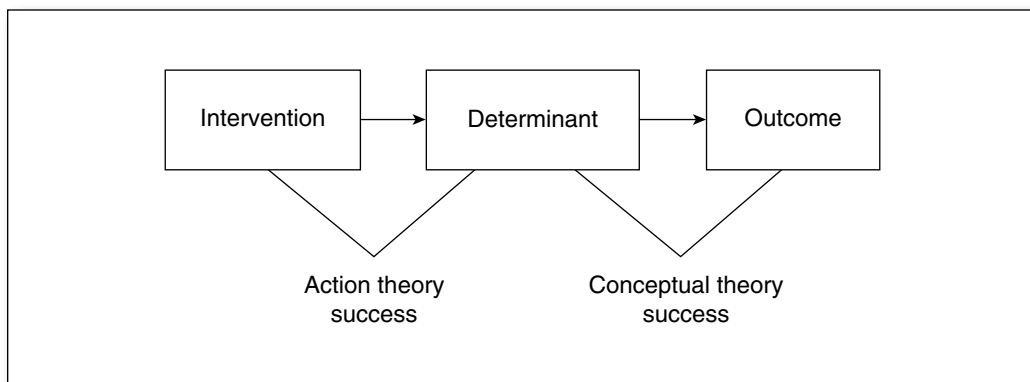
stakeholders tend to factor viability into their theory of an intervention, generally speaking, they tend to propose interventions with higher viability than do researchers. As demonstrated in this study, both outreach workers and clients in the housing complex setting were more receptive to the education/nonsmoking signage intervention than they were to CHA.

Action Theory Success and Conceptual Theory Success

In program theory literature, *action theory success* is the successful impact on a determinant by an intervention; otherwise, the intervention experiences action theory failure. *Conceptual theory success* occurs when a determinant successfully affects an outcome; otherwise, the intervention experiences conceptual theory failure (Chen, 1990, 2005). The differences between these two types of success are illustrated in Figure 14.5.

A successful intervention requires both action theory success and conceptual theory success. Generally speaking, a formal theory-based intervention tends to have an advantage in terms of conceptual theory success. The determinant in a social science theory-based intervention is usually a construct that has been intensively studied and has proved to be a powerful force for modifying human behaviors. Self-efficacy is such a determinant; it has proved to be an effective mechanism for changing human behaviors. The evaluation in this case study supports such an assertion. For example, self-efficacy was found to be effective in reducing exposure to tobacco smoke. But a formal theory-based intervention

Figure 14.5 Action Theory Success and Conceptual Theory Success



tends to be weak at the action theory level. That is, the intervention may not be potent enough to affect a determinant in the real world. As shown in this study, CHA failed to activate a self-efficacy change in this community setting. As researchers seek to develop formal theory-based interventions, they must learn how to develop robust interventions that can successfully affect a theoretical construct in a real-world setting.

Stakeholder theory-based interventions tend to experience more action theory success. A determinant in a stakeholder theory-based intervention usually reflects stakeholders' experience—that is, it has shown itself to be attainable through real-world intervention. In this case study, for example, the determinant is an environmental change expressed as the number of nonsmoking signage displays. This determinant is practical and easy for stakeholders to understand, and it is easier for stakeholders to organize intervention activities to affect signage displays than to affect self-efficacy. If a stakeholder theory-based intervention has weaknesses, they are more likely to be in the area of conceptual theory success. Determinants proposed by stakeholders may not be forceful enough to change human behavior, because these determinants do not benefit from ongoing study and refinement as do those supported by formal theory. Even so, this case study is encouraging: It demonstrates that a stakeholder theory-based intervention can have both action theory success and conceptual theory success. The results of this evaluation demonstrate the merit of conducting studies of strategies that would strengthen the conceptual theory success of stakeholder theory-based interventions.

LESSONS LEARNED FROM THE CASE STUDY

Formal theory and stakeholder theory are two major intervention bases. The widely held assumption has been that formal theory-based interventions are superior to stakeholder theory-based interventions. Still, few studies have empirically examined the relative merits of interventions based on the two theory types. The evaluation in the case study discussed in this chapter contributes to filling the gap in this important area by contrasting the relative merits of a formal theory-based intervention (CHA) and a stakeholder theory-based intervention (education/nonsmoking signage intervention) in a community setting.

CHA is based on well-studied social cognition theory and is supported by prior evidence. In terms of theory explicitness, CHA intervention also has advantages, as its underpinning theory is discussed thoroughly in the literature. In terms of implementation, however, the CHA intervention was much more resource- and labor-intensive than was the education/nonsmoking signage

intervention. Neither implementers nor clients were receptive to the CHA intervention. In addition, CHA was found to be ineffective in this real-world evaluation because it did not lead to an increase of self-efficacy, identified as the determinant in the causal chain of the program model.

One argument might be that the intervention's failure was due to implementation failure, perhaps due to unqualified implementers or implementers who did not spend enough time with clients. Yet even if this were true, stakeholders usually include implementation difficulties in their conception of the limitations of an intervention. When judging an intervention's merits, stakeholders go beyond efficacy to effectiveness. For example, stakeholders consider an intervention unwieldy if it requires hiring expensive implementers, if it exceeds the capacity of staff to reach and serve clients, or if it demands large changes in organizational structure or service routine.

This study is not alone in reporting challenges in implementing formal theory/evidence-based interventions in a real-world setting. For example, the Stanford Chronic Disease Self-Management Program (CDSMP), a formal theory/evidence-based intervention, requires a core group of health professional and lay leaders to take full responsibility for program activities, including preparing, scheduling, and leading courses and recruiting, registering, and following up with participants. Community-based organizations have faced challenges with implementing the program because they lack capacity or cannot maintain ongoing capacity (Freeman, Kadiyala, Bell, & Martin, 2008).

A common way that stakeholders cope with implementation difficulties is to modify a formal theory/evidence-based intervention to fit a specific situation or need, adapting or even reinventing it. For example, recipients of HIV-prevention funding were required to implement evidence-based interventions such as VOICES/VOCES and MPower (Veniegas, Kao, & Rosales, 2009). Fidelity to the implementation protocol was emphasized by the funding agency, but the community-based organizations in the study considerably modified or redesigned key characteristics of the evidence-based interventions. Changes included modifying the number and duration of sessions, adding extra elements, and altering content and delivery methods. The adaptations and reinventions were carried out during pre-implementation, implementation, and maintenance phases. With such adaptations and reinventions, a question arises as to how relevant the evidence provided by a formal theory/evidence-based intervention is to real-world application.

That said, the education/nonsmoking signage intervention has disadvantages. Principally, no prior evidence demonstrates its efficacy or effectiveness. Its theory is implicit and undocumented. For example, no previous studies supported this

education/nonsmoking signage intervention. Evaluators had to spend time and effort assisting stakeholders in clarifying their theory. Yet the evaluation found merit in this intervention in its real-world setting. It was found to be more viable than the CHA intervention in terms of implementers' and clients' acceptance. It was also relatively easy to implement and intuitively appealing. It increased the number of nonsmoking signage displays and reduced exposure to secondhand tobacco smoke, as predicted by the causal chain in the program model. Normally, stakeholder theory-based interventions tend to experience action theory success but be relatively weak in terms of conceptual theory success. This was not the case in this study. The theory of the education/nonsmoking signage intervention was well grounded in stakeholders' experience and observation, and this source in real-world experience may have enhanced its conceptual theory success.

The data indicate that the most influential variable was smoking status; that is, smokers at home were the biggest source of tobacco smoke pollution at home. This program only provided referral information to smokers for getting smoking-cessation treatment. Obviously, this was not sufficient to address the problem. It would be fruitful for future anti-secondhand smoke prevention programs to have stronger linkages or integration with smoking-cessation programs.

Furthermore, resource constraints and stakeholders' preference limited this evaluation to a two-group design: intervention and comparison. Formal theory- and stakeholder theory-based interventions were differentiated within the intervention group after implementation. This is not an ideal design for a formal comparison of two types of intervention. Formal and systematic comparisons of the two types of intervention would benefit from an application of stronger evaluation designs. For example, a three-group design using a formal theory group, a stakeholder theory group, and a comparison group would be desirable for comparing the two types of intervention in a rigorous way.

This case study provides some evidence for the merits of stakeholder theory and stakeholder theory-based interventions that have been neglected in the literature. It also provides some support for the arguments made by the integrated evaluation perspective, as discussed in Chapter 1, that is, for the need for evaluators to synthetically integrate the dynamic nature of an intervention program in a community and stakeholders' views and practices with existing scientific methods. The new perspective recognizes the potential merits of stakeholders' theories and interventions and implies the following future directions for intervention development and evaluation priority:

1. *Increase the study and evaluation of stakeholder theory-based interventions.* The new perspective stresses that intervention science should systematically

study, evaluate, aggregate, and disseminate stakeholder theory-based interventions just as researchers do with formal theory-based interventions. Funding agencies should provide resources to encourage the systematic study and evaluation of stakeholder theory-based interventions. Promising stakeholder theory-based interventions should have a chance to be rigorously studied and to become evidence-based interventions similar to formal theory-based interventions. Promising stakeholder theory-based interventions could be identified by using the integrative viability model (Chen, 2010; Chen & Garbe, 2011) or the Systematic Screening and Assessment Method (Leviton, Khan, & Dawkins, 2010).

2. *Expand the scope of formal theory-based interventions so as to address viability issues.* The new perspective recognizes, and the case study demonstrates, the contributions made by formal theory-based interventions. But for them to develop, the new perspective suggests a different route to their evaluation. Advocates of formal theory-based interventions have focused on providing evidence of efficacy but have neglected practical issues. Such a narrow focus means that generating stakeholder interest and application to real-world settings is difficult—unless, of course, funders mandate the use of formal theory-based interventions. The new perspective argues that evaluations of a formal theory-based intervention should address viability and transferability issues at the beginning, thereby ensuring that the intervention has support from stakeholders and has a good chance to prosper in a community (Chen, 2010).

3. *Integrate formal theory and stakeholder theory.* As indicated earlier in this chapter, both formal theory-based interventions and stakeholder theory-based interventions have their merits and demerits. Accordingly, an interchange between them may be fruitful. For example, researchers could learn from stakeholders about formulating a theory and an intervention with high viability. Likewise, stakeholders could learn from researchers how to enhance the sophistication of their theory and intervention or strengthen the rigor of the underlying evidence. Moreover, this collaboration could foster the development of theories and interventions interesting to both researchers and stakeholders. A major barrier to successful collaboration is that researchers and stakeholders do not have common interests or agendas. For example, stakeholders often view current formal theory as too abstract to apply to their practice, while researchers frequently regard stakeholder theory as too trivial to be worthy of investigation. Perhaps collaborative efforts could help to develop a new kind of middle-range intervention theory to which both researchers and stakeholders could relate, thereby narrowing the gap between science and service.

QUESTIONS FOR REFLECTION

1. What is the difference between formal theory-based interventions and stakeholder theory-based interventions? Give examples of each type of intervention.
2. Why are formal theory-based interventions the go-to intervention for researchers? Explain.
3. Discuss why stakeholders are often not interested in formal theory-based interventions?
4. Do you consider the experimentation evaluation approach or the holistic effectuality evaluation to be more appropriate for evaluating a stakeholder theory-based program? Why?
5. Discuss why the assumption that formal theory-based interventions supported by evidence of efficacy are superior to stakeholder theory-based interventions may or may not be incorrect.
6. What is the purpose of the process evaluation design in an outcome evaluation? Explain.
7. List the strengths and limitations of formal theory-based interventions.
8. List the strengths and limitations of stakeholder theory-based interventions.
9. Is evidence of efficacy better than real-world effectiveness? Why or why not?
10. Compare/contrast action theory success and conceptual theory success. Give examples.
11. Why do formal theory-based interventions tend to experience action theory success, while stakeholder theory-based interventions tend to experience conceptual theory success?
12. The author proposes three general strategies for mainstreaming stakeholder theory and stakeholder theory-based interventions. Discuss potential opportunities and hurdles for promoting these three strategies to the academic community and practice community.

CHAPTER 15



Evaluation and Dissemination

Top-Down Approach Versus Bottom-Up Approach

One of the important purposes of program evaluation is to disseminate evidence-based interventions. Evaluation is used to provide credible evidence to answer the question “Does an intervention program work?” If credible evidence indicates that the intervention works, decision makers, funding agencies, and scholars can use that information to promote and disseminate the intervention for application in various communities. On the surface, this seems to be a straightforward process, but in fact it is not. For example, what counts as credible evidence? Who are the audiences for evaluation results? What kinds of evaluation should be used? What is the stakeholders’ role in all of this? In spite of its importance, the transition from evaluation to dissemination is an understudied and underdiscussed area in program evaluation. This chapter will first discuss the traditional top-down approach to transitioning from evaluation to dissemination and its limitations. After that, the integrated evaluation perspective’s multifaceted view of what constitutes credible evidence of program success will be introduced, and the bottom-up approach will be proposed as an alternative for addressing dissemination-related issues.

THE TOP-DOWN APPROACH TO TRANSITIONING FROM EVALUATION TO DISSEMINATION

As discussed in Chapter 11, traditionally, two types of evaluation are identified as essential to determine whether an intervention should be disseminated: efficacy evaluation and effectiveness evaluation (Flay, 1986; Flay et al., 2005; Kellam & Langevin, 2003). An

efficacy evaluation (or efficacy study) uses an RCT to evaluate an intervention or treatment in an ideal, highly controlled clinical research setting to achieve the highest possible internal validity. In contrast, effectiveness evaluation assesses the effectiveness of an intervention in real-world conditions.

The top-down approach to the transition from evaluation to dissemination, used in health promotion and social betterment programs, is to sequentially conduct an efficacy evaluation and then an effectiveness evaluation of an intervention. According to this approach, a new intervention must first undergo an efficacy evaluation (using an RCT) to maximize the internal validity of the assessment (Flay et al., 2005). After the intervention's efficaciousness is determined, a real-world effectiveness evaluation is applied to address the evaluation's external validity (generalizability). Only after an intervention proves efficacious in a controlled setting and effective in the real world is it deemed suitable for dissemination.

The top-down approach of social betterment and health promotion programs is a very brief version of the long path from research to dissemination used in biomedical research. Biomedical research begins with testing a new drug on animals, proceeding through four phases of clinical trials with multiple RCTs to rigorously determine the drug's safety, efficacy, effectiveness, and dissemination. This process is outlined by the US Food and Drug Administration (FDA, 1992). Phases I and II address dose and safety issues. After a drug is deemed safe for testing in humans, it is formally evaluated through a sequence of evaluations, namely efficacy evaluations (Phase III) and effectiveness evaluations (Phase IV). It often takes multiple years or decades for a new drug to reach the dissemination stage. This approach is recognized by many scientists as the "gold standard" of the scientific method.

The adapted version of the top-down approach used in health promotion and social betterment programs only uses these three steps in the transition process: efficacy evaluation, effectiveness evaluation, and dissemination. This cycle is far shorter than that used in biomedical research. However, since it is related to the gold standard of the scientific method, many funding agencies, researchers, and health promotion and social betterment program evaluators are attracted to it.

LESSONS LEARNED FROM APPLYING THE TOP-DOWN APPROACH TO PROGRAM EVALUATION

The top-down approach's major contributions to program evaluation include the establishment of stringent standards and concrete methods that enhance the rigor of outcome evaluation. This approach has improved program

evaluation's scientific reputation by proposing the use of RCTs to provide the strongest source of credible evidence of an intervention's effects. In fact, the term *evidence-based interventions* or *evidence-based practices* is often limited only to RCT-evaluated interventions. Yet in applying the top-down approach, health promotion/social betterment programs have experienced the following problems:

1. *An evidence-based intervention does not imply that the intervention is likely to be effective in the real world.* Because of a heavy emphasis on internal validity, evaluations following the top-down approach tend to neglect external validity. Usually, efficacy evaluations do not follow effectiveness evaluations (Green & Glasgow, 2006). In other words, evidence-based interventions in their current form usually provide little evidence of real-world generalizability. When advocates promote or disseminate to stakeholders an evidence-based intervention, those advocates must assume that the intervention is likely to be effective in the real world. The truth of this assumption is unknown due to a lack of effectiveness evaluations.

In fact, the few sequential evaluations that have been conducted have shown mixed results. On the one hand, interventions such as Coordinated Approach to Child Health (Luepker et al., 1996) have proved to be successful in both research and real-world settings. On the other hand, the evaluation results of Reconnecting Youth (Hallfors et al., 2006) sounded real-world alarm bells. According to the initial efficacy evaluation of this program (Eggert, Thompson, Herting, Nicholas, & Dicker, 1994), the intervention was found among other things to decrease involvement with hard drug use and increase students' grade point average. But in a subsequent effectiveness evaluation, not only did the intervention not have desirable effects on drug control problems and school performance, but on measures of peer bonding, high-risk peer bonding, and socially desirable weekend activities, program adolescents actually had worse outcomes than did those in the control group (Hallfors et al.). The authors argued that the harmful effects might have resulted from the iatrogenic effects of grouping high-risk youths in a real-world setting.

Dissemination of evidence-based interventions is potentially precarious if that intervention's real-world generalizability is unknown. As demonstrated in the Reconnecting Youth study, an evidence-based intervention might not only be ineffective in the real world but even harmful.

2. *The conditions under which an evidence-based intervention is designed and administered often do not resemble or are irrelevant to real-world operations.* When assessing an intervention's effect, one efficacy evaluation purpose is to create an ideal and controlled setting to maximize internal validity. The

problem, however, is that creation of an ideal and controlled setting for health promotion/social betterment programs decreases the intervention's real-world relevancy. The use of different types of implementers illustrates the issue. To ensure an intervention's appropriate delivery, highly qualified and enthusiastic implementers often provide the efficacy evaluation. But as demonstrated in the Open Airways for Schools (OAS) program, these implementers tend to little resemble their real-world counterparts. OAS is a school-based health education program devised to enhance third- to sixth-graders' ability to manage their asthma on a daily basis. Designers intended that the program would facilitate the ability of parents and children to work with their clinicians to manage the disease. An efficacy evaluation in fact found that the intervention was efficacious in attaining program goals (D. Evans, Clark, & Feldman, 1987). However, the team delivering the OAS evaluation comprised, among others, doctoral-level sociologists and educators with master's degrees in public health and social sciences (Bruzzese, Markman, Appel, & Webber, 2001). These highly qualified OAS efficacy evaluation counselors only remotely resembled real-world OAS counselors. In the real world, volunteer educators deliver this program. These are parents with limited OAS training or limited training in behavioral change generally. Many community-based organizations cannot afford to hire highly qualified counselors. Thus, stakeholders may view efficacy evaluation evidence as irrelevant to real-world situations.

Originally, the use of highly qualified counselors in efficacy evaluations may have been intended to enhance internal validity. But the manipulation of research conditions and setting also creates an environment that boosts intervention effects. Highly qualified counselors usually are more knowledgeable and skilled than real-world counselors in changing clients' beliefs or behaviors. From the real-world standpoint then, highly qualified and enthusiastic counselors in an efficacy evaluation may artificially inflate the intervention's desirable effects.

Note that different types of counselors are only one of many contextual factors that could be manipulated in efficacy evaluation. Other contextual factors include monetary or other incentives provided to participants. Moreover, such factors as supervision, coordination, and the recruitment process could also be manipulated to boost internal validity—and intervention effects. Unfortunately, the literature on evidence-based interventions usually emphasizes methodological rigor and statistical precision. It largely ignores the extent of contextual factor manipulation that might contribute to intervention effects (Chen, 1990, 2005). This might give audiences who are unaware of the manipulation of external factors a false impression or hope that application of these evidence-based interventions will result in similarly desirable real-world outcomes.

3. *Evidence-based interventions do not address the practical or service issues that are highly relevant to stakeholders.* Stakeholders are responsible for contacting and delivering services to clients on a day-to-day basis. Stakeholders have concerns about practical issues such as whether an intervention will be attractive to real-world clients, whether the intervention will be suitable for ordinary implementers to deliver, and whether a typical community organization is capable of managing the intervention. Although from a researcher's standpoint these practical issues may be regarded as trivial, to stakeholders they are crucial. Yet efficacy evaluations often do not address these issues. This is another example of why, because most evidence-based interventions do not adequately address practical issues, stakeholders generally do not find them useful (Wandersman et al., 2008).

The need for evaluation to address stakeholders' views and concerns is clearly reflected in the four standards of program evaluation: utility, feasibility, propriety, and accuracy (Sanders & Joint Committee on Standards for Educational Evaluation, 1994). Current forms of evidence-based interventions mainly focus on the fourth standard and do not adequately address the first three standards, which are concerned primarily with practical and service issues. This difference in focus results in a huge gap between intervention research and real-world practice (Chen, 2005; Green & Glasgow, 2006). Today more and more practitioners, decision makers, and consumers find that traditional scientific evaluation results tend not to be useful to the everyday issues about which those groups are concerned (Wandersman et al., 2008).

4. *Evidence-based interventions are difficult to implement in the real world.* As discussed previously, stakeholders' low enthusiasm for evidence-based interventions does not imply they do not put evidence-based interventions into practice. Because of funding agency requirements, community-based organizations often implement evidence-based interventions. Just as often, however, when community-based organizations attempt to implement evidence-based interventions, they encounter real-world challenges. The difficulties are illustrated in the real-world implementation of the National Cooperative Inner-City Asthma Study (NCICAS). NCICAS was a randomized clinical trial of an intervention that used trained master's level social workers to make frequent contact with families to deliver asthma counseling and to deal with the families' psychosocial needs (R. Evans et al., 1999). The RCT, as is typical in efficacy evaluation, provided participants with

- monetary and child care incentives,
- highly committed counselors,
- food/refreshments,

- frequent contacts between participants and counselors, and
- counseling sessions during regular office hours.

The evaluation found that the NCICAS intervention was efficacious in reducing asthma morbidity among inner-city children.

The Inner-City Asthma Intervention (ICAI) wanted to implement the NCICAS intervention in a real-world setting (Kattan, 2006; Williams & Redd, 2006). However, ICAI experiences demonstrate the difficulty of delivering a scientific evidence-based intervention with high fidelity in the real world; in fact, many adaptations or changes were required. For example, ICAI social workers were unable to contact and meet with families as frequently as in NCICAS. The great majority of ICAI counseling sessions were held in the evenings, on weekends, or both in an attempt to promote continued participation in the intervention. Due to budget constraints, ICAI could not provide monetary and child care incentives to clients or food/refreshments to increase the enjoyment of session attendance, as had NCICAS. ICAI also found retaining social workers difficult. Because of these implementation difficulties, at the end only 25% of the children in ICAI completed the entire intervention.

As shown by the ICAI experience, a common mechanism that stakeholders use to cope with implementation difficulties is to modify an evidence-based intervention to fit a specific situation or need. Modification often means adapting or even reinventing an evidence-based intervention. For example, recipients of HIV-prevention funding were required to implement evidence-based interventions such as VOICES/VOCES and Mpower (Veniegas et al., 2009). The grant announcement required community-based organizations receiving the funds to stress fidelity in implementing evidence-based interventions, but the study found that these organizations either considerably modified key characteristics or even redesigned the evidenced interventions. Changes included modification of the number and duration of sessions, addition of extra elements, and modification of the intervention content and delivery methods. The adaptations and reinventions were carried out during pre-implementation, implementation, and maintenance phases.

Due to adaptation or reinvention, an evidence-based intervention and its real-world counterpart tend to differ substantially. And two versions of an intervention can create confusion in understanding, communicating, or disseminating intervention. Such questions arise as, How relevant is an evidence-based intervention's evidence to its real-world counterpart? Can the adapted or reinvented intervention even be called an evidence-based intervention? If evidence-based interventions require adaptation and reinvention for real-world applicability, what are the real purposes or benefits of evidence-based interventions? Given

such questions, we need to ask, Is the top-down approach the only path to ensuring the scientific quality of health promotion/social betterment programs?

Interventions with credible evidence usually mean those evaluated by RCTs or other experimental methods. The academic community and many funding agencies are keen to disseminate evidence-based interventions to communities, for use in solving problems. Unfortunately, the transitions of many programs have not been successful. Stakeholders are not interested in these evidence-based interventions; there is a huge gap between the academic and practice community regarding what is desirable in interventions.

To a large degree, this disconnect results from the fact that a vast number of intervention programs, although efficacious, have no applicability to the real world (Chinman et al., 2005; Glasgow et al., 2003; Green & Glasgow, 2006; Rotheram-Borus & Duan, 2003; Wandersman, 2003; Wandersman et al., 2008). On the other hand, stakeholders have adopted or adapted popular interventions despite the absence of strong evidence supporting their effectiveness. For example, DARE remains the most widely disseminated substance abuse prevention program (D. C. Gottfredson & Gottfredson, 2002; G. D. Gottfredson & Gottfredson, 2001), despite evaluations showing that it does not demonstrate positive results (Clayton, Cattarello, & Johnstone, 1996; Lynam et al., 1999).

The above case histories indicate that fundamental assumptions underlying the top-down approach do not always fit well with health promotion/social betterment program evaluation. And the problem cannot be easily resolved by technical strategies such as capacity building, technical assistance, or translational research. To address credible evidence and dissemination issues, program evaluation may need a more comprehensive perspective and model. This chapter will use the integrative evaluation perspective to develop an integrative cogency model for evaluative evidence and a bottom-up approach to address the transition from evaluation to dissemination.

INTEGRATIVE COGENCY MODEL: THE INTEGRATED EVALUATION PERSPECTIVE

The integrated evaluation perspective argues that the framework of credible evidence provided by the top-down approach does not take into consideration stakeholders' views, practices, and needs. This may be one of the major reasons stakeholders are not interested in evidence-based interventions provided under the top-down approach. To address the problem, the integrated evaluation perspective proposes the integrative cogency model for evaluative evidence. The

concept of *cogency* is defined in this book as the extent that the evidence of an evaluation's conclusions is clear, logical, and convincing. In a real-world evaluation, stakeholders hope evaluations can provide evidence in multiple areas. As discussed in Chapter 10, the experimentation evaluation approach proposes to use the Campbellian validity typology to conceptualize evaluation evidence in two areas: internal validity and external validity. The ideas of internal and external validity provide a starting point for evaluators addressing issues of evidence. However, current thinking is that program evaluation needs to address evidence issues beyond the distinction between internal and external validity. For example, Spencer and colleagues (2013) proposed a conceptual framework with five types of evidence to support evidence-based practice: effectiveness, reach, feasibility, sustainability, and transferability.

Working in a similar direction, this chapter proposes the integrative cogency model, which argues that evaluative evidence of an intervention should consist of three components:

1. *Effectuality*. The extent to which an evaluation provides evidence that an intervention affects specific goals/outcomes
2. *Viability*. The extent to which an evaluation provides evidence that an intervention is viable in the real world
3. *Transferability*. The extent to which an intervention's effectuality and viability are transferable from research to real-world settings or from one real-world setting to another

Evaluators must provide cogent evidence in these three areas. Accordingly, the integrative cogency model consists of three types of cogency: effectual cogency, viable cogency, and transferable cogency.

Obviously, the development of the integrative cogency model benefited greatly from the Campbellian validity typology. However, since the Campbellian validity typology was developed for and is applicable to experimental research, as explained in Chapter 11, the integrated evaluation perspective proposes the term *cogency*, instead of *validity* to avoid confusion.

Effectual Cogency

The type of cogency that is most familiar to evaluators is effectual cogency. Effectual cogency is how credible the evidence is that shows whether an intervention has desirable effects on program outcomes. The reason for effectual cogency's

popularity is that it directly relates to goal attainment. *Goal attainment* refers to whether an intervention can achieve the targets set for it. It is important to note that there are two kinds of effectual cogency. The first kind is an intervention's pure independent effects as discussed in Chapter 10. The second kind is an intervention's joint effects in the real world, as discussed in Chapter 11. In communicating on effectual cogency issues with stakeholders, evaluators need to clearly indicate which type of effectual cogency they are referring to.

Viable Cogency

Before explaining the concept of viable cogency, it is important to discuss goal attainment and system integration. Goal attainment is important to stakeholders, but they are equally or even more interested in system integration. *System integration* means the extent to which an intervention is compatible or in synergy with other components such as mission, culture, manpower, structure, and capacity in a real-world organization or community in a system.

Note also that although goal attainment and system integration are related outcomes attributable to an intervention, they do not necessarily go hand in hand. An efficacious or effective intervention does not necessarily integrate well with a community-based organization, or vice versa. For example, a school-based intervention found to be efficacious may have problems in system integration if its implementation requires schools to overhaul their existing schedules. On the other hand, an intervention integrated well with a community-based organization is not necessarily effective. DARE is a popular substance abuse prevention program among schools, but evaluations have shown its ineffectiveness (Lynam et al., 1999). A successful intervention program does well in both goal attainment and system integration.

To address system integration issues, the integrative cogency model proposes a new concept of viable cogency. Viable cogency is the extent to which an intervention succeeds or thrives in the real world. The concept of viable cogency is derived from various literatures.* Here, *viable cogency* refers to stakeholders'

*The research from which the concept of viable cogency is derived includes that dealing with diffusion of innovation (E. M. Rogers, 2003; Zaltman & Duncan, 1977), capacity of community organizations (CDC, 1999; Chen, 2001; Chinman et al., 2005; Spoth & Greenberg, 2005), cultural appropriateness (Bamberger, Rugh, & Mabry, 2012; Miller & Shinn, 2005), and stakeholders' views and needs (Cunningham-Sabo et al., 2007; Glasgow et al., 2003; Sandler et al., 2005; Wandersman, Valois, et al., 1996; Wandersman, Duffy, et al., 2008; Weiner, Helfrich, Savitz, & Swiger, 2007).

views, based on their experience with an intervention program, regarding whether it is practical, affordable, suitable, evaluable, and helpful. More specifically, viable cogency refers to whether ordinary practitioners—rather than research staff—can implement an intervention program adequately, and whether the intervention program is suitable for coordination or management by a service delivery organization such as a community clinic or a community-based organization. Additional questions are whether the intervention program (1) is affordable, (2) can recruit ordinary clients without paying them to participate, (3) has a clear rationale for its structure and linkages to expected outcomes, and (4) is regarded by actual clients and other stakeholders as helpful in alleviating clients' problems or in enhancing their well-being. In this context, *helpful* is defined as whether stakeholders notice or experience progress in alleviating or resolving a problem.

Viable cogency can only be attained if evaluators invite stakeholders in a sufficiently inclusive manner to provide input. For example, if evaluators neglect the viewpoint of a major stakeholder group, such as implementers or clients, evaluators may reach an improper conclusion about the viability of an intervention program. If evaluators cannot provide a safe environment in which stakeholders can express their views, the stakeholders may not be willing to provide candid information about a program's viability. Strategies for dealing with such potential problems include bringing in representatives from major stakeholder groups to plan the evaluation, triangulating qualitative and quantitative data, ensuring effective communication between evaluators and stakeholders about the purpose of the evaluation and how evaluation data will be used, or a combination thereof.

In the real world, stakeholders organize and implement an intervention program for the purpose of serving clients. Thus, stakeholders have real viability concerns. In many situations, even though a program lacks strong evidence of its effectuality, if it appears to be an innovative, viable program, stakeholders adopt it. For example, the Resolving Conflict Creatively Program (RCCP) is a comprehensive, K–12 school violence-prevention program (Aber, Brown, Chaudry, Jones, & Samples, 1996). It seeks to change the classroom, peer group, and school contexts in which children learn how to resolve conflict. The program was first developed in 1985 by a collaboration of Educators for Social Responsibility (a Cambridge, Massachusetts, not-for-profit organization) and the New York City Board of Education. The program started in three elementary schools in Brooklyn, New York, and appealed greatly to school administrators, teachers, students, and parents (Selfridge, 2004). The program empowered teachers, students, and parents to contribute to a positive and safe school environment, and it was regarded as highly viable. Because of its perceived viability and support

from the New York Board of Education, within several years the program grew to serve thousands of children in over 90 New York City schools. It was also widely adopted by schools in several other states. All of this dissemination occurred several years before the program was formally evaluated for its effectiveness (Aber et al.).

Transferable Cogency

The third type of cogency considered in the integrative cogency model is transferable cogency. Transferable cogency refers to the ability of intervention effectiveness or viability to transfer from research to real-world settings or from one real-world setting to another. This type of cogency is related to the concept of external validity, but has some differences. Under Campbell and Stanley's (1963) validity typology, external validity asks the following question: "To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" According to this definition, external validity involves an endless quest for confirmation of an intervention's universal worth. This open-ended quest for generalizability may be appropriate for research, but it is impossible for evaluation to adequately address. Thus, the concept of transferable cogency is developed for evaluation purposes. In terms of transferability, transferable cogency has a boundary—the real world. Evaluators are capable of addressing transferability issues within this boundary. Furthermore, transferable cogency expands the scope of external validity from effectuality to both effectuality and viability; in this way, it better reflects stakeholders' interests (Chen, 2010).

EVALUATION APPROACHES RELATED TO THE INTEGRATIVE COGENCY MODEL

Effectuality Evaluation

This book has intensively discussed two approaches for addressing effectual cogency. Chapter 10 discussed how to apply the experimental evaluation approach to assess an intervention's independent, pure effects. Chapter 11 discussed how to apply the holistic effectuality evaluation approach to assess an intervention's real-world effects. Factors to consider when determining which approach to apply first will be discussed later.

Viability Evaluation

With viable cogency as one of its components, the integrative cogency model points to another evaluation type. This type of evaluation—identified here as *viability evaluation*—assesses the extent to which an intervention program is viable in the real world. More specifically, it evaluates whether likely implementers can reasonably implement an intervention, whether the intervention can recruit participants, whether the implementing organization is capable of coordinating intervention-related activities, whether the intervention is viewed as affordable, and whether likely clients and other stakeholders feel the intervention has real-world value. Viability evaluation requires the use of mixed methods research (Greene & Caracelli, 1997; Tashakkori & Teddlie, 2003). On the one hand, viability evaluation relies on quantitative methods to collect data to monitor progress on recruitment, retention, and outcomes. On the other hand, it requires an in-depth understanding of stakeholders' views on and experiences with an intervention program; to understand such views and experiences, qualitative methods such as focus groups and in-depth interviews are highly useful.

Transferability Evaluation

Transferability evaluation assesses the extent to which an evaluation's findings of a program's effectiveness can be generalized from a research setting to a real-world setting or from one real-world setting to another. Transferable cogency is achieved when an evaluation itself provides sufficient contextual factors for an intervention to be deemed effective in real-world applications. Potential users can interpret the information on the effectiveness of the intervention in light of the setting's contextual factors. Users can thereby assess its potential to be generalized to their own populations and settings and decide whether to apply the intervention in their communities.

Exhibited generalization can be achieved through the action model/change model framework as described by the theory-driven approach (Chen, 1990, 2005). The exhibited generalization approach (Chen et al., 2014) provides in-depth information on contextual factors and causal mechanisms embedded in an intervention program's viability and effectuality so that potential users can make an informed decision about whether an intervention program could be transferred to their community.

Stakeholders sometimes have a particular real-world target population or setting to which they want to generalize evaluation results. This is known as

targeted generalization; it is the extent to which evaluation results can be generalized to a specific target population and real-world setting. Targeted generalization is achieved through methods such as the sampling approach (Shadish et al., 2002), Cronbach's (1982) UTOS approach, or the dimension test approach (Chen, 1990). Thus, through exhibited or targeted generalization, external validity is potentially achievable in a program evaluation.

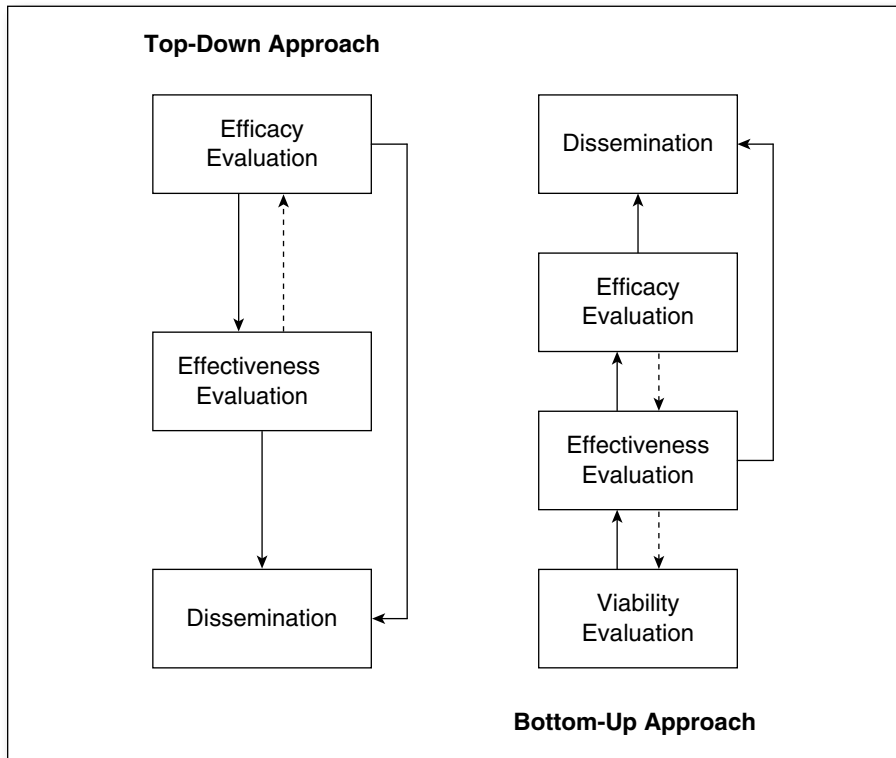
THE BOTTOM-UP APPROACH TO TRANSITIONING FROM EVALUATION TO DISSEMINATION

Based upon the integrative cogency model and its associated evaluation approaches, the integrated evaluation perspective proposes the bottom-up approach to transition from evaluation to dissemination.

The Bottom-Up Approach

Under the effectual, viable, and transferable framework of the integrative cogency model, the bottom-up approach provides another route to the achievement of cogency (Chen, 2010; Chen and Garbe, 2011). The steps of this approach occur in reverse order from those of the top-down approach. To maximize viable cogency (i.e., Is the intervention practical, affordable, suitable, and helpful?), the evaluation starts with viability evaluation. If the real-world intervention is found viable, an *effectiveness evaluation* provides systematic and objective evidence of the intervention's effectiveness in the real world. If necessary, the effectiveness evaluation could also address whether the effectiveness is generalizable to other real-world settings. After the intervention is found to be viable, effective, and generalizable to real-world settings, an *efficacy evaluation*, using methods such as RCTs, is conducted to assess rigorously a causal relationship between intervention and outcome. Although this approach is particularly relevant to social betterment/health promotion programs, it has not been formally or systematically discussed in the literature.

Figure 15.1 highlights the differences between the bottom-up and top-down approaches. Solid-line arrows indicate the sequence of evaluations leading to dissemination; dotted-line arrows indicate where feedback is given to facilitate improvement. As indicated in the figure, the top-down approach proceeds from efficacy evaluation, to effectiveness evaluation, and then to viability evaluation, while the bottom-up approach starts from viability evaluation and goes through effectiveness evaluation to efficacy evaluation. Researchers and scientists are

Figure 15.1 Top-Down Approach Versus Bottom-Up Approach

the driving force of top-down approach interventions. Stakeholders are the driving force of bottom-up approach interventions, although to assist in designing their interventions, they may use information and guidance provided by researchers and scientists. The top-down approach takes a science-to-service route, while the bottom-up approach takes a service-to-science route.

The Bottom-Up Approach and Social Betterment/Health Promotion Programs

The theory and methodology of the bottom-up approach fit well with the political, organizational, and community environments typically surrounding social betterment/health promotion programs. As mentioned previously, a social betterment/health promotion program is usually started as a real-world

program in a response to a pressing problem (Bamberger et al., 2012; Chen, 2005; Patton, 2008; Rossi et al., 2004; Weiss, 1998). The literature might provide some guidance, but stakeholders themselves are usually responsible for developing the program, and they usually do so under time constraints. Given such conditions, stakeholders are eager for an evaluation to provide field evidence regarding whether the intervention program could successfully reach and recruit participants; whether it has broad support from the community; whether an organization can smoothly run it; and whether implementers, clients, or other key stakeholders feel the intervention is helpful in the field. Such information is useful to stakeholders for strengthening the same program in the future or for designing and implementing similar programs in other settings. Under these conditions, conducting a viability evaluation to address viable cogency issues makes sense, especially to demonstrate whether the program is practical and helpful in the real world.

This does not mean, however, that stakeholders are uninterested in knowing precisely whether a causal relationship exists between the intervention and outcomes. Rather, they may feel that this type of time-consuming and resource-intensive evaluation does not fit their immediate evaluation needs. In their eyes, such efficacy evaluation can wait until a future round of evaluation. A program deemed “viable” then undergoes an effectiveness evaluation to assess its effectiveness and whether it is generalizable to a real-world setting. If an intervention is found to be viable, effective, and generalizable to the real world, efficacy evaluation would be conducted to develop the scientific knowledge that would allow an assessment of the causal relationship between the intervention and outcomes.

Needle exchange programs (NEPs) provide one example of the bottom-up approach. Such programs prevent HIV transmission by providing injection drug users (IDUs) with new, sterile syringes in exchange for their used syringes, thus discouraging needle sharing. A local Junkiebond (Junkie League), a not-for-profit organization formed by and for illicit drug users, introduced the first NEP in Amsterdam, the Netherlands, in 1984 (Van den Hoek, van Haastrecht, & Coutinho, 1989). The intervention program quickly became widespread. Many community-based organizations were attracted to NEPs because the intervention shares their harm-reduction philosophy and is relatively easy to apply in communities. In addition to sterile syringes, many NEPs provide condoms and clean sterile equipment or paraphernalia (e.g., cottons, cookers, bleach) that facilitate safer injection. The drug users themselves support the program’s viability because view it as nonthreatening and accessible. Thus since the mid-1980s, a number of developed and developing countries have introduced NEPs as a core component of HIV prevention targeting IDUs (Stimson, 1996).

Evaluators and researchers have conducted NEP effectiveness evaluations. Several studies have consistently found that NEPs are effective in reducing injection-related risk behaviors (Bluthenthal, Kral, Erringer, & Edlin, 1998; Des Jarlais et al., 1996; Lurie, 1997; Vlahov et al., 1997) as well as reducing the incidence of HIV (Des Jarlais & Friedman, 1998; Hagan, Des Jarlais, Friedman, Purchase, & Alter, 1995). Meta-analyses of a large number of effectiveness evaluations (Cross, Saunders, & Bartelli, 1998; Ksobiech, 2003) concluded that NEPs contribute to a reduction in needle sharing. Recently, RCTs have been used to evaluate NEPs, providing the strongest evidence of NEP effectiveness (Masson et al., 2007).

Types of Intervention for the Bottom-Up Approach

The bottom-up approach can be widely applied in different types of interventions. Since this approach emphasizes viability issues, stakeholders' interventions are particularly appropriate for it. The NEPs discussed above are a good example. If researchers are interested in ensuring their interventions' viability before going into more rigorous evaluations, they can apply this approach to evaluate their interventions. This approach is also appropriate for interventions jointly developed by stakeholders and researchers.

THE CURRENT VERSION OF EVIDENCE-BASED INTERVENTIONS: LIMITATIONS AND STRATEGIES TO ADDRESS THEM

The conceptual framework of the integrative cogency model is also useful for evaluators who want to examine critically the concept of credible evidence and the controversies related to this concept. With interest growing in evidence-based interventions, it is crucial for the evaluation community to adequately and systematically address issues around what constitutes credible evidence. Unfortunately, to date, little consensus exists among evaluators regarding this issue (Donaldson et al., 2009). To better understand the questions involved, this section briefly introduces the history of the evidence-based intervention movement and discusses the major disagreements among evaluators regarding credible evidence. This section then applies the integrative cogency model to provide a balanced view of what credible evidence is and to assist the evaluation community in effectively communicating and addressing issues related to credible evidence.

The evidence-based intervention movement has grown in popularity across disciplines. The movement originally started in medicine in the 1990s

(Atkins, Fink, & Slutsky, 2005; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). According to Sackett and colleagues, evidence-based medicine is the “conscious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients” (p. 71). Evidence-based medicine intends to overcome the limitations of the “old style” of clinical practice, which is based upon unsystematic observations from clinical experience. The evidence-based medicine movement has spread to public health (Kohatsu, Robinson, & Torner, 2004; McGinnis & Foege, 2000) and to many social and behavioral disciplines under the general label of *evidence-based interventions*.

The evidence-based intervention movement relies on RCTs to maximize a study’s internal validity. Proponents of the movement have argued that only evidence produced by RCTs is credible (Nutbeam, 1999; Speller et al., 1997; Stephenson & Imrie, 1998; Tilford, 2000). The position has been challenged for its relevancy to health promotion programs and to public health in general (Britton, Thorogood, Coombes, & Lewando-Hundt, 1998; World Health Organization [WHO], 1998). Here is an example that illustrates the debate: In 2003, the then US secretary of education proposed placing a priority on scientifically based evaluation methods for funding competitive grants. The proposal identified randomized experimental methods as the best method for determining what constitutes scientifically based evaluation for assessing intervention effects. Interestingly enough, as discussed in Chapter 11, in responding to the challenge the evaluation community was divided. The American Evaluation Association (AEA) leadership issued a policy statement opposing efforts to prioritize RCTs in education evaluation-funding competitions, arguing that nonrandomized methods are capable of generating understanding of causality (Donaldson & Christie, 2005). This does not mean, however, that evaluators have reached consensus on this issue. A group of senior members of the American Evaluation Association opposed the AEA’s statement and issued a response known as “The Not AEA statement” (Donaldson & Christie, 2005). They argued that among interventions in many other areas of public policy—including health and medicine, mental health, criminal justice, employment, and welfare—RCTs have been essential to understanding what works, what does not work, and what is harmful. Attempts to draw conclusions about intervention effects based upon nonrandomized trials have often led to misleading results in those fields, and, this group concluded, there is no reason to expect the same is not true in the social and education fields.

For many years, the evidence debates have concentrated on internal validity issues. This book attempts to open another front for future discussions by asking

whether only evidence of internal validity should be deemed credible. This question is crucial. If internal validity is just one portion of what makes evidence credible, then current evidence-based intervention movements may not be building their arguments on a solid foundation. The bottom line is whether credible evidence is a one-dimensional or multidimensional concept. It is interesting to note that even researchers/scientists in medicine have argued that credible evidence should be a multidimensional concept. For example, Sackett and colleagues (1996) pointed out that evidence-based medicine means an integration of individual clinical expertise with the best available external evidence from systematic research. They argued that using evidence-based medicine does not mean applying the best external evidence in slavish, cookbook fashion. Similarly, Atkins et al. (2005) and Haynes (1999) have noted that when policy makers assess scientific evidence of a medical intervention, they ask not only “Can it work?” but also “Will it work?” and “Is it worth it?”

The integrative cogency model posits that credible evidence of social betterment/health promotion programs is a multidimensional concept. As described above, credible evidence is a set of three related types of evidence: viable cogency (viability), effectual cogency (effectiveness), and transferable cogency (transferability). Under this model, the current view of credible evidence—that it is equivalent to efficacy, that is, evidence provided by RCTs—is very narrow. It is important to point out that efficacy represents only one type of effectuality: an intervention’s pure independent effects. In terms of effectuality, as demonstrated in Chapter 11, it is also important to know an intervention’s real-world (joint) effects.

Furthermore, efficacy or effectuality is not a stand-alone or context-free concept. Rather, it should be viewed or discussed as a reference point for viability and transferability. The problem of equating evidence of efficacy or effectuality with a totality of evidence, without reference to transferability, is illustrated as follows: Effectiveness of a social betterment/health promotion intervention is contingent on contextual factors such as the types of implementers, the implementing organizations, and clients. If the context of an intervention is changed, the effectiveness of the intervention is also likely to change. For example, an innovative intervention is evaluated in a controlled setting. Clients are paid to ensure their participation and retention. Furthermore, intensively trained, highly paid, and highly motivated research staff are used to implement the intervention to ensure its fidelity. The evaluation provides strong evidence of intervention efficacy in the controlled setting. Since the current evidence-based approach counts only efficacy as evidence, researchers will classify the intervention as evidence based.

If, however, transferability is considered, the picture of credible evidence is altered. In the real world, clients are not paid for participation, and the intervention is typically implemented by the staffs of community-based organizations rather than by research staff. Because of the major difference between the controlled setting and the real world, the efficacy of the controlled setting is likely not found in the real world. From the standpoint of stakeholders, the controlled-setting efficacy may be an artificial effect, relevant or transferable only to that artificial situation. Accordingly, if the evidence-based intervention movement counts only efficacy as an evidentiary criterion, it might promote an intervention with strong but “artificial” evidence of effectiveness for real-world use. An even worse scenario is that a funding agency may require community-based organizations to adopt the “artificial” evidence-based intervention as a condition for receiving funds. In this case, community-based organizations may be forced to implement an ineffective intervention in communities.

Similar arguments can be applied to the question of whether evidence of viability should be factored in as part of credible evidence. An effective intervention in controlled settings is not necessarily a viable one in the real world. And when stakeholders are not able or not willing to implement an unviable intervention, this kind of intervention—no matter how strong the evidence of efficacy produced in controlled settings—is useless to them. Again, if the evidence-based intervention movement uses only efficacy as credible evidence, it may mistakenly promote those interventions that have little chance of survival in the real world. A push to implement an effective intervention without evidence of real-world viability would not only be a waste of valuable resources—it would also be unscientific. The integrative cogency model argues that credible evidence must include viability, effectuality, and transferability. The integrative cogency model and the bottom-up approach may aid advocates of the evidence-based intervention movement to move from current narrow views of evidence to a well-based, credible evidence model (Urban, Hargraves, & Trochim, 2014).

THE INTEGRATED EVALUATION PERSPECTIVE ON CONCURRENT COGENCY APPROACHES

Under the conceptual framework of the integrated evaluation perspective, concurrent cogency approaches contemplate dealing with multiple cogency issues in a single evaluation. A concurrent approach has important implications for program evaluation. For programs dealing with such problems as high school dropouts, unemployment, low access to health care, unsafe sex, crime, and so on, evaluators may be asked to conduct one evaluation rather than multiple

outcome evaluations. Outcome evaluation is time-consuming—the turnaround time for an outcome evaluation of any such program could easily be a few years. A long turnaround time is one of the major reasons why stakeholders ask for one outcome evaluation rather than multiple outcome evaluations for the same program. Because the simultaneous maximizing of effectual, viable, and transferable cogency when conducting an evaluation using a concurrent cogency approach is impossible, evaluators must choose one of the following approaches to deal with cogency issues in an evaluation: focusing on effectual cogency, focusing on viable cogency, or optimizing cogency. These approaches and their applications are briefly discussed below.

Focusing on Effectual Cogency

If stakeholders are more interested in effectuality than other issues, evaluators can focus an evaluation on addressing effectual cogency. In doing so, evaluators must understand whether stakeholders are interested in assessing an intervention's pure independent effects or real-world (joint) effects, as well as taking into account the funds provided for evaluation. If stakeholders are interested in assessing an intervention's pure independent effects, evaluators must select the experimentation evaluation approach, as discussed in Chapter 10. If funds are sufficient and no ethical or administration concerns arise, then evaluators could select RCT or other experimental methods to provide the strongest evidence of an intervention's pure independent effects.

However, if stakeholders are more interested in assessing an intervention's real-world effects, evaluators can follow the real-world effectuality evaluation approach, discussed in Chapter 11, to design an evaluation.

Focusing on Viable Cogency

If stakeholders need information about whether a program is practical or doable in the real world or whether real-world organizations, implementers, and clients favor the program, the viable cogency-focused approach is an excellent choice. Evaluators could apply a viability evaluation for this purpose. Viability evaluation requires data on implementers', clients', and other stakeholders' views on and experiences with an intervention program. RCTs or randomized experiments are not appropriate methods for this kind of evaluation. Instead, for viability evaluation, mixed methods (i.e., qualitative and quantitative) are particularly appropriate (Greene & Caracelli, 1997; Tashakkori

& Teddlie, 2003). Qualitative methods such as intensive interviews, focus groups, and case studies and quantitative methods such as surveys are highly useful for this purpose.

Optimizing Approach

If stakeholders prefer that an evaluation provide in one evaluation evidence of two or three types of cogency (e.g., viable, effectual, and transferable), the optimizing approach is an excellent choice. Because maximizing multiple types of validity in one evaluation is impossible, evaluators turn to *optimizing*, which focuses on finding a good-enough solution to the requirement that an evaluation address multiple validities (Chen, 1988, 1990; Chen & Rossi, 1983b). The effectiveness evaluation is an example of applying the optimizing approach to achieve multiple validities. The optimizing evaluation approach usually requires the use of mixed methods (Greene & Caracelli, 1997; Tashakkori & Teddlie, 2003).

The integrative cogency model argues that when evaluators apply these concurrent approaches, they avoid the limitations of a dogmatic view that effectual cogency is always superior to other types of cogency. As previously discussed, effectual cogency is not necessarily the prime priority in every evaluation. Whether a particular evaluation approach is appropriate is contingent on contextual factors such as stakeholders' interests and the nature of a program. For example, if stakeholders were keen to demonstrate that the programs are practical and survivable in the real world, an effectual cogency-focused approach would be the wrong approach. Instead, a viable cogency-focused evaluation can provide evidence of the viability of a program in the real world. Similarly, if stakeholders want evidence on the real-world effectiveness, generalizing capability, and viability of an intervention, optimizing is a better choice than trying to maximize internal validity. To select an appropriate evaluation type that ensures the relevance and usefulness of the evaluation, evaluators must understand stakeholders' needs and the nature of the intervention.

THE USEFULNESS OF THE BOTTOM-UP APPROACH AND THE INTEGRATIVE COGENCY MODEL

The advantages of the bottom-up approach and the integrative cogency model to program evaluation include the following:

1. *Ensure the intervention's usefulness to stakeholders and avoid wasting money.* The traditional top-down approach usually begins with an expensive and time-consuming efficacy evaluation to assess an innovative intervention. After many dollars are spent, it might be found that the efficacious intervention is very difficult to implement in the real world, not of interest to stakeholders, or not real-world effective. As a consequence, the evidence-based intervention may not be useful. This approach can waste money and resources.

By contrast, the bottom-up approach starts with a viability evaluation, which assesses the viability of an intervention as proposed by researchers or stakeholders. The use of viability evaluation assures that an intervention has a good chance to survive in the real world before an expensive effectiveness or efficacy evaluation, or both, are undertaken. Because interventions with low viability are screened out to begin with, this approach can save funding agencies considerable resources. The bottom-up approach encourages funding agencies to fund many viability evaluations in communities to accumulate scientific knowledge on viable interventions and then to select highly viable interventions for further rigorous studies.

2. *Provide an opportunity to revise and improve an intervention in the real world before its finalization.* One top-down approach limitation is that it finalizes the intervention protocol or package before or during efficacy evaluation—the protocol is not supposed to change after the evaluation. When an intervention protocol is finalized at such an early stage, the intervention is barred from the benefits of feedback from the experience of real-world implementation or stakeholders' input. This approach seriously restricts an intervention's generalizability to the real world.

By contrast, the bottom-up approach affords an opportunity to improve an intervention during the viability evaluation. One viability evaluation purpose is to enhance the quality and usefulness of an intervention from the standpoint of organizational and community dynamics. Intervention protocols developed from stakeholder input and implementation experience increase the intervention's real-world relevancy and contributions.

3. *Provide a balanced view based on credible evidence.* The integrative cogency model subsumes credible evidence from three components: viability, effectuality, and transferability. Under this model, evidence on intervention effectuality (internal validity) is not a stand-alone or context-free concept. Rather, it should be viewed or discussed with reference to viability and transferability.

This model may aid certain advocates of evidence-based intervention to move from valuing only single-dimension evidence (effectuality) to embracing a more balanced, multidimensional evidence model (viability, effectuality, and transferability).

4. *Provide a fresh contingency view on methods.* As discussed in Chapter 1 of this book, evaluators have debated intensely whether RCTs or other methods are the best research methods for evaluation. The integrative cogency model provides a contingency view that may be useful in reconciling differences between the opposing camps. The contingency view argues that different methods are useful for addressing different validity issues. For example, it recognizes the power of RCTs in enhancing effectual cogency, but unlike the top-down approach, argues against a wide application of RCTs in evaluation. Instead, RCTs must be applied carefully and only for those interventions already assessed by viability evaluation and by effectiveness evaluation so as to avoid wasting money and other valuable resources.

Similarly, in addressing viable and transferable cogency issues and advocating for their greater application in evaluation, the integrative cogency model recognizes the essential value of qualitative methods, which address viability and transferability issues, and the contributions of quantitative methods, which enhance effectual cogency in evaluation. Because the contingency view emphasizes strengths and limitations of different methods within different evaluation contexts, this view might be more acceptable to both quantitative and qualitative camps. Moreover, it might reconcile the differences between these two camps by narrowing the gap or identifying common ground.

QUESTIONS FOR REFLECTION

1. What is the top-down approach for the transition from evaluation to dissemination? Why do funding agencies and academia prefer this approach?
2. Why do you think evidence-based interventions are generally limited to those evaluated via RCTs?
3. Compare and contrast interventions that operate in a biological system versus in a social system.

4. The top-down approach is regarded as the gold standard of scientific research in biological intervention. Should the same standard apply to social betterment/health promotion programs? Explain your reasoning.
5. Discuss lessons learned from applying the top-down approach to health promotion/ social betterment programs.
6. Why aren't RCTs effective when the evaluation is intended to address practical or service issues relevant to stakeholders?
7. Discuss relationships between the integrated evaluation perspective and the integrative cogency model.
8. Compare and contrast the typology of internal and external validity and the typology of viable, effectual, and transferable cogency. Why might one want to shift from a validity typology to a cogency typology?
9. What is the bottom-up approach? How does it differ from the top-down approach?
10. Which approach—top-down or bottom-up—is more useful for health promotion/ social betterment programs under various conditions? Explain.
11. Discuss two limitations of the current version of evidence-based interventions and provide the potential solutions for each.
12. Why is it important for an evaluator, when conducting an effectual cogency evaluation, to understand whether a stakeholder is interested in assessing an intervention's pure independent effects or real-world (joint) effects?
13. Would it be appropriate to utilize the viable cogency approach if a stakeholder was interested in assessing an intervention's pure independent effects? Why or why not?
14. Discuss potential barriers an evaluator might encounter when conducting an evaluation with the top-down approach and the bottom-up approach. And what are the benefits of each approach?

References

- Aber, J. L., Brown, J. L., Chaudry, N., Jones, S. M., & Samples, F. (1996). The evaluation of the Resolving Conflict Creatively Program: An overview. *American Journal of Preventive Medicine*, 12(5 Suppl.), 82–90.
- Ajzen, I., & Fishbein, M. (Eds.). (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Alkin, M. C. (2013). *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed.). Thousand Oaks, CA: SAGE.
- American Evaluation Association. (2004). Guiding principles for evaluators. <http://www.eval.org/p/cm/ld/fid=51>
- Atkins, D., Fink, K., & Slutsky, J. (2005). Better information for better health care: The Evidence-based Practice Center program and the Agency for Healthcare Research and Quality. *Annals of Internal Medicine*, 142(12 Pt. 2), 1035–1041.
- Bamberger, J. M., Rugh, J., & Mabry, L. S. (2012). *RealWorld evaluation: Working under time, data, and political constraints* (2nd ed.). Thousand Oaks, CA: SAGE.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Bartholomew, L. K., Parcel, G. S., Kok, G., & Gottlieb, N. H. (2001). *Intervention mapping: Designing theory- and evidence-based health promotion programs*. Mountain View, CA: Mayfield.
- Berk, R. A., Lenihan, K. J., & Rossi, P. H. (1980). Crime and poverty: Some experimental evidence for ex-offenders. *American Sociological Review*, 45, 766–786.
- Bertalanffy, L. von. (1968). *General system theory: Foundations, development, applications* (Rev. ed.). New York, NY: George Braziller.
- Bickman, L. (1987). The function of program theory. In L. Bickman (Ed.), *Using program theory in evaluation*. San Francisco, CA: Jossey-Bass.
- Bickman, L. (Ed.). (1990). *Advances in program theory*. San Francisco, CA: Jossey-Bass.
- Bickman, L. (1996). The application of program theory to a managed mental health care evaluation. *Evaluation and Program Planning*, 19(2), 111–119.
- Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, 15, 253–268.
- Bluthenthal, R. N., Kral, A. H., Erringer, E. A., & Edlin, B. R. (1998). Use of an illegal syringe exchange and injection-related risk behaviors among street-recruited injection drug users in Oakland, California, 1992 to 1995. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 18(5), 505–511.
- Britton, A., Thorogood, M., Coombes, Y., & Lewando-Hundt, G. (1998). Search for evidence of effective health promotion: Quantitative outcome evaluation with qualitative process evaluation is best. *BMJ*, 316(7132), 703–704.

- Brunk, S. E., & Goeppinger, J. (1990). Process evaluation: Assessing re-invention of community-based interventions. *Evaluation & the Health Professions*, 13(2), 186–203.
- Bruzzese, J. M., Markman, L. B., Appel, D., & Webber, M. (2001). An evaluation of Open Airways for Schools: Using college students as instructors. *The Journal of Asthma*, 38(4), 337–342.
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Centers for Disease Control and Prevention (CDC). (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report (MMWR)*, 48(RR 11), 1–40.
- Chelimsky, E. (1987). The politics of program evaluation. In D. S. Cordray, H. S. Bloom, & R. J. Light (Eds.), *New directions in program evaluation: Vol. 34. Evaluation practice in review* (pp. 5–21). San Francisco, CA: Jossey-Bass.
- Chelimsky, E. (1997). The political environment of evaluation and what it means for the development of the field. In E. Chelimsky and W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 53–68). Thousand Oaks, CA: SAGE.
- Chen, H.-T. (1988). Validity in evaluation research: A critical assessment of current issues. *Policy and Politics*, 16(1), 1–16.
- Chen, H.-T. (1990). *Theory-driven evaluations*. Thousand Oaks, CA: SAGE.
- Chen, H.-T. (1996). A comprehensive typology for program evaluation. *Evaluation Practice* 17(2), 121–130.
- Chen, H.-T. (1997). Normative evaluation of an anti-drug abuse program. *Evaluation and Program Planning*, 20(2), 195–204.
- Chen, H.-T. (2001). Development of a national evaluation system to evaluate CDC-funded health department HIV prevention programs. *American Journal of Evaluation*, 22(1), 55–70.
- Chen, H.-T. (2003). Theory-driven approach for facilitation of planning health promotion or other programs. *Canadian Journal of Program Evaluation*, 18(2), 91–113.
- Chen, H.-T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: SAGE.
- Chen, H.-T. (2010). The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning*, 33(3), 205–214.
- Chen, H.-T. (2012a). The roots and growth of theory-driven evaluation: Assessing viability, effectuality, and transferability. In M. Alkin (Ed.), *Evaluation roots*, 113–129. Thousand Oaks, CA: SAGE.
- Chen, H.-T. (2012b). Theory-driven evaluation: Conceptual framework, application, and advancement. In R. Strobl, O. Lobermeier, & W. Heitmeyer (Eds.), *Evaluation von Programmen und Projekten für eine demokratische Kultur [Evaluation of programs and projects for a democratic culture]* (pp. 17–40). Wiesbaden, Germany: Springer VS.
- Chen, H.-T., Donaldson, S. I., & Mark, M. M. (Eds.). (2011). *Advancing validity in outcome evaluation: Theory and practice*. New Directions for Evaluation. Volume 130. San Francisco, CA: Jossey-Bass.
- Chen, H.-T., & Garbe, P. (2011). Assessing program outcomes from the bottom-up approach: An innovative perspective to outcome evaluation. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130; pp. 93–106). San Francisco, CA: Jossey-Bass.

- Chen, H.-T., Grimley, D. M., Waithaka, Y., Aban, I. B., Hu, J., & Bachmann, L. H. (2008). A process evaluation of the implementation of a computer-based, health provider-delivered HIV prevention intervention for HIV-positive men who have sex with men in the primary care setting. *AIDS Care*, 20(1), 51–60.
- Chen, H.-T., Liang, S., Liao, Q., Wang, S., Schumacher, J., Creger, T., . . . Vermund, S. (2007). HIV voluntary counseling and testing among injection drug users in south China: A study of a non-government organization based program. *AIDS and Behavior*, 11(5), 778–788.
- Chen, H.-T., & Mark, M. (1996). Assessing the needs of inner city youth: Beyond needs identification and prioritization. *Children and Youth Services Review*, 18(3), 1–31.
- Chen, H.-T., Quane, J., & Garland, T. N. (1988). Evaluating an antismoking program. *Evaluation & the Health Professions*, 11(4), 441–464.
- Chen, H.-T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces*, 59(1), 106–122.
- Chen, H.-T., & Rossi, P. H. (1983a). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7(3), 283–302.
- Chen, H.-T., & Rossi, P. H. (1983b). The theory-driven approach to validity. *Evaluation and Program Planning*, 10, 95–103.
- Chen, H.-T., & Turner, N. (2012). Formal theory vs. stakeholder theory: New insights from a tobacco-focused prevention program evaluation. *American Journal of Evaluation*, 33(3), 395–413.
- Chen, H.-T., Wang, J. C. S., & Lin, L. H. (1997). Evaluating the process and outcome of a garbage reduction program in Taiwan. *Evaluation Review*, 21(1), 27–42.
- Chen, H.-T., Yip, F., Lavonas, E. J., Iqbal, S., Turner, N., Cobb, B., & Garbe, P. (2014). *Using the exhibited generalization approach to evaluate a carbon monoxide ordinance*. Unpublished manuscript.
- Chinman, M., Hannah, G., Wandersman, A., Ebener, P., Hunter, S. B., Imm, P., & Sheldon, J. (2005). Developing a community science research agenda for building community capacity for effective preventive interventions. *American Journal of Community Psychology*, 35(3–4), 143–157.
- Christie, C. A. (Ed.). (2003). *The theory-practice relationship in evaluation* (New Directions for Evaluation, No. 97). San Francisco, CA: Jossey-Bass.
- Chun, Y. H., & Rainey, H. G. (2005). Goal ambiguity and organizational performance in U.S. federal agencies. *Journal of Public Administration Research and Theory*, 15(4), 529–557.
- Clapp, J. D., & Early, T. J. (1999). A qualitative exploratory study of substance abuse prevention outcomes in a heterogeneous prevention system. *Journal of Drug Education*, 29(3), 217–233.
- Clayton, R. R., Cattarello, A. M., & Johnstone, B. M. (1996). The effectiveness of Drug Abuse Resistance Education (project DARE): 5-year follow-up results. *Preventive Medicine*, 25(3), 307–318.
- Clotfelter, C. T., & Hahn, J. C. (1978). Assessing the national 55 mph speed limit. *Policy Sciences*, 9, 281–294.
- Connell, J. P., Kubisch, A. C., Schorr, L. B., & Weiss, C. H. (1995). *New approaches to evaluating community initiatives: Concepts, methods and contexts*. Washington, DC: Aspen Institute.

- Conybeare, J. A. (1980). Evaluation of automobile safety regulations: The case of compulsory seat belt legislation in Australia. *Policy Sciences*, 12, 27–39.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cornell, C. E., Littleton, M. A., Greene, P. G., Pulley, L., Brownstein, J. N., Sanderson, B. K., . . . Raczynski, J. M. (2009). A Community Health Advisor Program to reduce cardiovascular risk among rural African-American women. *Health Education Research*, 24(4), 622–633.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226.
- Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: SAGE.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C. . . . Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass.
- Cross, J. E., Saunders, C. M., & Bartelli, D. (1998). The effectiveness of educational and needle exchange programs: A meta-analysis of HIV preventional strategies for injection drug users. *Quality & Quantity*, 32(2), 165–180.
- Cunningham-Sabo, L., Carpenter, W. R., Peterson, J. C., Anderson, L. A., Helfrich, C. D., & Davis, S. M. (2007). Utilization of prevention research: Searching for evidence. *American Journal of Preventive Medicine*, 33(1 Suppl.), S9–S20.
- Davis, M., Baranowski, T., Resnicow, K., Baranowski, J., Doyle, C., Smith, M., . . . Yaroch, A. (2000). Gimme 5 fruit and vegetables for fun and health: Process evaluation. *Health Education Behavior*, 27(2), 167–176.
- Des Jarlais, D. C., & Friedman, S. R. (1998). Fifteen years of research on preventing HIV infection among injecting drug users: What we have learned, what we have not learned, what we have done, what we have not done. *Public Health Reports*, 113(Suppl. 1), 182–188.
- Des Jarlais, D. C., Marmor, M., Paone, D., Titus, S., Shi, Q., Perlis, T. . . . Friedman, S. R. (1996). HIV incidence among injecting drug users in New York City syringe-exchange programmes. *Lancet*, 348(9033), 987–991.
- Donaldson, S. I. (2001). Mediator and moderator analysis in program development. In S. Sussman (Ed.), *Handbook of program development for health behavior research & practice* (pp. 470–496). Thousand Oaks, CA: SAGE.
- Donaldson, S. I. (2003). Theory-driven program evaluation in the new millennium. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems* (pp. 109–141). Mahwah, NJ: Lawrence Erlbaum.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York, NY: Lawrence Erlbaum.
- Donaldson, S. I., & Christie, C. A. (2005). The 2004 Claremont Debate: Lipsey versus Scriven. Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation*, 3, 60–67.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2008). *What counts as credible evidence in applied and evaluation practice?* Thousand Oaks, CA: SAGE.

- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2015). *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2nd ed.). Thousand Oaks, CA: SAGE
- Donaldson, S. I., Graham, J. W., & Hansen, W. B. (1994). Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug abuse prevention interventions. *Journal of Behavioral Medicine*, 17(2), 195–215.
- Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*, 70(11), 35–36.
- Durkheim, É. (2014). *The division of labor in society*. (Trans. W. D. Halls et al.), New York, NY: The Free Press. (Originally published in France in 1893)
- Dutton, D. G. (1986). The outcomes of court-mandated treatment for wife assault: A quasi-experimental evaluation. *Violence and Victims*, 1(3), 163–175.
- Eggert, L. L., Thompson, E. A., Herting, J. R., Nicholas, L. J., & Dicker, B. G. (1994). Preventing adolescent drug abuse and high school dropout through an intensive school-based social network development program. *American Journal of Health Promotion*, 8(3), 202–215.
- Ethridge, M. E., & Percy, S. L. (1993). A new kind of public policy encounters disappointing results: Implementing Learnfare in Wisconsin. *Public Administration Review*, 54(4), 340–347.
- Evans, D., Clark, N. M., Feldman, C. H. (1987). School health education programs for asthma. *Clinical Reviews in Allergy*, 5(3), 207–212.
- Evans, R., III, Gergen, P. J., Mitchell, H., Kattan, M., Kercsmar, C., Crain, E. . . . Wedner, H. J. (1999). A randomized clinical trial to reduce asthma morbidity among inner-city children: Results of the National Cooperative Inner-City Asthma Study. *The Journal of Pediatrics*, 135(3), 332–338.
- Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (Eds.). (2015). *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability* (2nd ed.). Thousand Oaks, CA: SAGE.
- Food and Drug Administration (FDA). (1992). *Guideline for the clinical evaluation of analgesic drugs*. Rockville, MD: FDA.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15(5), 451–474.
- Flay, B. R., Biglan, A., Boruch, R. F., González Castro, F., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175.
- Fleury, J., Keller, C., Perez, A., & Lee, S. M. (2009). The role of lay health advisors in cardiovascular risk reduction: A review. *American Journal of Community Psychology*, 44(1–2), 28–42.
- Foster, G. H. (1973). *Traditional societies and technical change*. New York, NY: Harper & Row.
- Freeman, J. D., Kadiyala, S., Bell, J. F., & Martin, D. P. (2008). The causal effect of health insurance on utilization and outcomes in adults: A systematic review of US studies. *Medical Care*, 46(10), 1023–1032.
- Fulbright-Anderson, K., Kubisch, A. C., & Connell, J. P. (Eds.). (1998). *New approaches to evaluating community initiatives. Vol. 2: Theory, measurement and analysis*. Washington, DC: Aspen Institute.
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey Bass.

- Gargani, J., & Donaldson, S. L. (2011). What works for whom, where, why, for what, and when? Using evaluation evidence to take action in local contexts. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130; pp. 17–30). San Francisco, CA: Jossey-Bass.
- Gettleman, L., & Winkleby, M. A. (2000). Using focus groups to develop a heart disease prevention program for ethnically diverse, low-income women. *Journal of Community Health, 25*(6), 439–453.
- Glantz, K., Carbone, E., & Song, V. (1999). Formative research for developing targeted skin cancer prevention for children in multi-ethnic Hawaii. *Health Education Research: Theory and Practice, 14*(2), 155–166.
- Glasgow, R. E., Lando, H., Hollis, J., McRae, S. G., & La Chance, P. A. (1993). A stop-smoking telephone help line that nobody called. *American Journal of Public Health, 83*(2), 252–253.
- Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93*(8), 1261–1267.
- Glassman, M., Lacson, R., Collins, B., Hill, C., & Wan, C. (2002). Lessons learned from the first year of implementation of the Centers for Disease Control and Prevention's standardized evaluation system for HIV prevention programs. *AIDS Education and Prevention, 14*(3A), 49–58.
- Goodman, R. M., Wandersman, A., Chinman, M., Imm, P., & Morrissey, E. (1996). An ecological approach of community-based interventions for prevention and health promotion: Approaches to measuring community coalitions. *American Journal of Community Psychology, 24*(1), 33–61.
- Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school based prevention programs: Results from a national survey. *Journal of Research on Crime and Delinquency, 39*(1), 3–35.
- Gottfredson, G. D., & Gottfredson, D. C. (2001). What schools do to prevent problem behavior and promote safe environments. *Journal of Education and Psychological Consultation, 12*, 313–344.
- Gottlieb, N. H., Lovato, C. Y., Weinstein, R., Green, L. W., & Eriksen, M. P. (1992). The implementation of a restrictive worksite smoking policy in a large decentralized organization. *Health Education Quarterly, 19*(1), 77–100.
- Gowdy, E. A., & Freeman, E. (1993). Program supervision: Facilitating staff participation in program analysis, planning and change. *Administration in Social Work, 17*(3), 59–79.
- Green, L. W., & Glasgow, R. E. (2006). Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Evaluation & the Health Professions, 29*(1), 126–153.
- Green, L. W., & Kreuter, M. W. (1991). *Health promotion planning: An educational and environmental approach* (2nd ed.). Mountain View, CA: Mayfield.
- Greene, J. (2009). Evidence as “proof” and evidence as “inkling.” In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 153–167). Thousand Oaks, CA: SAGE.
- Greene, J. C. (2011). The construct(ion) of validity as argument In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation:*

- Theory and practice* (New Directions for Evaluation, No. 130, pp. 81–92). San Francisco, CA: Jossey-Bass.
- Greene, J., & Caracelli, V. J. (Eds.). (1997). *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (New Directions for Evaluation, Vol. 74). San Francisco, CA: Jossey-Bass.
- Greene, J., Lincoln, Y. S., Mathison, S., Mertens, D. M., & Ryan, K. (1998). Advantages and challenges of using inclusive evaluation approaches in evaluation practice. *American Journal of Evaluation*, 19(1), 101–122.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco, CA: Jossey-Bass.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth-generation evaluation*. Newbury Park, CA: SAGE.
- Hagan, H., Des Jarlais, D. C., Friedman, S. R., Purchase, D., & Alter, M. J. (1995). Reduced risk of hepatitis B and hepatitis C among injection drug users in the Tacoma syringe exchange program. *American Journal of Public Health*, 85(11), 1531–1537.
- Halfors, D., Cho, H., Sanchez, V., Khatapoush, S., Kim, H. M., & Bauer, D. (2006). Efficacy vs. effectiveness trial results of an indicated “model” substance abuse program: Implications for public health. *American Journal of Public Health*, 96(12), 2254–2259.
- Hansen, M. B., & Vedung, E. (2010). Theory-based stakeholder evaluation. *American Journal of Evaluation*, 31(3), 295–313.
- Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L.A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the Adolescent Alcohol Prevention Trial. *Journal of Studies on Alcohol*, 52(6), 568–579.
- Hawe, P., & Stickney, E. K. (1997). Developing the effectiveness of an intersectoral food policy coalition through formative evaluation. *Health Education Research*, 12(2), 213–225.
- Haynes, B. (1999). Can it work? Does it work? Is it worth it? The testing of health care interventions is evolving. *BMJ*, 319(7211), 652–653.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: SAGE.
- House, E. R. (2011). Conflict of interest and Campbellian validity. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130, pp. 69–80). San Francisco, CA: Jossey-Bass.
- Jones, J. H. (1993). *Bad blood: The Tuskegee syphilis experiment* (Rev. ed.). New York, NY: The Free Press. (Original edition published 1981)
- Julian, D., Jones, A., & Deyo, D. (1995). Open systems evaluation and the logic model: Program planning and evaluation tools. *Evaluation and Program Planning*, 18(4), 333–341.
- Julnes, G. (2011). Reframing validity in research and evaluation: A multidimensional, systematic model of valid inference. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130, pp. 55–68). San Francisco, CA: Jossey-Bass.
- Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28(2), 167–172.

- Kattan, M. (2006). Moving research into the community: Learning by experience. *Annals of Allergy, Asthma, & Immunology*, 97(1 Suppl. 1), S2–3.
- Kellam, S. G., & Langevin, D. J. (2003). A framework for understanding “evidence” in prevention research and programs. *Prevention Science*, 4(3), 137–153.
- Kohatsu, N. D., Robinson, J. G., & Torner, J. C. (2004). Evidence-based public health: An evolving concept. *American Journal of Preventive Medicine*, 27(5), 417–421.
- Kohler, C. L., Grimley, D., & Reynolds, K. D. (1999). Theoretical approaches guiding the development and implementation of health promotion programs. In J. M. Raczynski & R. J. DiClemente (Eds.), *Handbook of health promotion and disease prevention* (pp. 23–50). New York, NY: Kluwer Academic.
- Kristal, A. R., Glanz, K., Tilley, B. C., & Li, S. (2000). Mediating factors in dietary change: Understanding the impact of a worksite nutrition intervention. *Health Education & Behavior*, 27(1), 112–125.
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.) Thousand Oaks, CA: SAGE.
- Ksobiech, K. (2003). A meta-analysis of needle sharing, lending, and borrowing behaviors of needle exchange program attenders. *AIDS Education and Prevention*, 15(3), 257–268.
- Laub, C., Somera, D. M., Gowen, L. K., & Diaz, R. M. (1999). Targeting “risky” gender ideologies: Constructing a community-driven theory-based HIV prevention intervention for youth. *Health Education & Behavior*, 26(2), 185–199.
- Leviton, L. C., Khan, L. K., & Dawkins, N. (Eds.). (2010). *The systematic screening and assessment method: Finding innovations worth evaluating* (New Directions for Evaluation, Vol. 123). San Francisco, CA: Jossey-Bass.
- Luepker, R. V., Perry, C. L., McKinlay, S. M., Nader, P. R., Parcel, G. S., Stone, E. J., . . . Johnson, C. C. (1996). Outcomes of a field trial to improve children’s dietary patterns and physical activity. The Child and Adolescent Trial for Cardiovascular Health. CATCH collaborative group. *JAMA*, 275(10), 768–776.
- Lurie, P. (1997). Invited commentary: Le mystère de Montreal. *American Journal of Epidemiology*, 146(12), 1003–1006.
- Lynam, D. R., Milich, R., Zimmerman, R., Novak, S. P., Logan, T. K., Martin, C., . . . Clayton, R. (1999). Project DARE: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology*, 67(4), 590–593.
- Mark, M. M. (2003). Toward an integrative view of the theory and practice of program and policy. In S. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems* (pp. 183–204). Mahwah, NJ: Lawrence Erlbaum.
- Mark, M. M. (2011). New (and old) directions for validity concerning generalizability. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130, pp. 31–42). San Francisco, CA: Jossey-Bass.
- Mark, M. M., Hofmann, D. A., & Reichardt, C. S. (1992). Testing theories in theory-driven evaluation: (Tests of) moderation in all things. In H.-T. Chen & P. H. Rossi (Eds.), *Using theory to improve program and policy evaluations* (pp. 71–84). Westport, CT: Greenwood.
- Martin, M. Y., Kim, Y. I., Kratt, P., Litaker, M. S., Kohler, C. L., Schoenberger, Y. M., . . . Williams, O. D. (2011). Medication adherence among rural, low-income

- hypertensive adults: A randomized trial of a multimedia community-based intervention. *American Journal of Health Promotion*, 25(6), 372–378.
- Marx, R., Hirozawa, A. M., Chu, P. L., Bolan, G. A., & Katz, M. H. (1999). Linking clients from HIV antibody counseling and testing to prevention services. *Journal of Community Health*, 24(3), 201–214.
- Masson, C. L., Sorensen, J. L., Perlman, D. C., Shopshire, M. S., Delucchi, K. L., Chen, T., . . . Hall, S. M. (2007). Hospital- versus community-based syringe exchange: A randomized controlled trial. *AIDS Education and Prevention*, 19(2), 97–110.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3), 279–300.
- McGinnis, J. M., & Foege, W. (2000). Guide to community preventive services: Harnessing the science. *American Journal of Preventive Medicine*, 18(1 Suppl.), 1–2.
- McKillip, J. (1992). Research without control groups: A control construct design. In F. B. Bryant, J. Edwards, R. S. Tindale, E. J. Posavac, L. Heath, E. Henderson-King, & Y. Suarez-Balcazar (Eds.), *Methodological issues in applied social psychology* (pp. 159–175). New York, NY: Plenum.
- McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your program performance story. *Evaluation and Program Planning*, 22(1), 65–72.
- Meadows, D. H. (2008). *Thinking in systems: A primer*. White River Junction, VT: Chelsea Green.
- Mercier, C., Piat, M., Peladeau, N., & Dagenais, C. (2000). An application of theory driven evaluation to a drop-in youth center. *Evaluation Review*, 24(1), 73–91.
- Miller, R. L., & Shinn, M. (2005). Learning from communities: Overcoming difficulties in dissemination of prevention and promotion efforts. *American Journal of Community Psychology*, 35(3–4), 169–183.
- Miller, W. R., Toscova, R. T., Miller, J. H., & Sanchez, V. (2000). A theory-based motivational approach for reducing alcohol/drug problems in college. *Health Education & Behavior*, 27(6), 744–759.
- Nkwake, A. (2013). *Working with assumptions in international development program evaluation*. New York, NY: Springer.
- Nutbeam, D. (1999). Evaluating health promotion. *BMJ*, 318(7180), 404.
- O’Leary, A., Jemmott, L. S., Goodhart, F., & Gebelt, J. (1996). Effects of an institutional AIDS prevention intervention: Moderation by gender. *AIDS Education and Prevention*, 8(6), 516–528.
- Olsen, L. (1981). *Ohio District Eleven Adoption Project for Handicapped Children: Final evaluation report*. Cleveland, OH: Human Services Design Laboratory, School of Applied Social Sciences, Case Western Reserve University.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: SAGE.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.
- Pawson, R., & Tilly, N. (1997). *Realistic evaluation*. Thousand Oaks, CA: SAGE.
- Perrow, C. (1961). The analysis of goals in complex organizations. *American Sociological Review*, 26(6), 854–866.
- Posavac, E. J. (2011). *Program evaluation: Methods and case studies* (8th ed.). Boston, MA: Prentice-Hall.

- Quantock, C., & Beeynon, J. (1997). Evaluating an osteoporosis service using a focus group. *Nursing Standard*, 11(42), 45–47.
- Reeler, D. (2007). *A three-fold theory of social change and implications for practice, planning, monitoring and evaluation*. Woodstock, South Africa: Community Development Resource Association.
- Reichardt, C. S. (2011). Criticisms of and an alternative to the Shadish, Cook, and Campbell validity typology. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130, pp. 43–54). San Francisco, CA: Jossey-Bass.
- Renger, R., & Titcomb, A. (2002). Three-step approach to teaching logic models. *American Journal of Evaluation*, 23(4), 493–503.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
- Rogers, P. J., Hasci, T. A., Petrosino, A., & Huebner, T. A. (Eds.). (2000). *Program theory in evaluation: Challenges and opportunities* (New Directions for Evaluation, No. 87). San Francisco, CA: Jossey-Bass.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: SAGE.
- Rotheram-Borus, M. J., & Duan, N. (2003). Next generation of preventive interventions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(5), 518–526.
- Ryan, A., & Bohman, J. (1998). Systems theory in social science. In *Routledge Encyclopedia of Philosophy* (Version 1.0). London, England: Routledge.
- Sabatier, P. A. (Ed.). (1999). *Theories of the policy process*. Boulder, CO: Westview.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it isn't. *BMJ*, 312(7023), 71–72.
- Sanders, J. R., & The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: SAGE.
- Sandler, I., Balk, D., Jordan, J., Kennedy, C., Nadeau, J., & Shapiro, E. (2005). Bridging the gap between research and practice in bereavement: Report from the Center for the Advancement of Health. *Death Studies*, 29(2), 93–122.
- Schneider, M., Ituarte, P., & Stokols, D. (1993). Evaluation of a community bicycle helmet promotion campaign: What works and why. *American Journal of Health Promotion*, 7(4), 281–287.
- Scriven, M. (1967). *The methodology of evaluation perspectives of curriculum evaluation* (Vol. I, pp. 39–83). Chicago, IL: Rand McNally.
- Scriven, M. (1972). Pros and cons about goal-free evaluation. *The Journal of Educational Evaluation*, 3(4), 1–4.
- Scriven, M. (1991a). Beyond formative and summative evaluation. In G. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: A quarter century* (pp. 19–64). Chicago, IL: University of Chicago Press.
- Scriven, M. (1991b). [Pros] and cons about goal-free evaluation. *American Journal of Evaluation*, 12(1), 55–76.
- Scriven, M. (1997). Truth and objectivity in evaluation. In E. Chelimsky and W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 477–500). Thousand Oaks, CA: SAGE.

- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19, 57–70.
- Scriven, M. (2008). A summative evaluation of RCT methodology and an alternative approach to causal research. *Journal of MultiDisciplinary Evaluation*, 5(9), 11–24.
- Selfridge, J. (2004). The Resolving Conflict Creatively Program: How we know it works. *Theory Into Practice*, 43(1), 59–67.
- Shadish, W. R., Jr. (2011). The truth about validity. In H.-T. Chen, S. L. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (New Directions for Evaluation, No. 130, pp. 107–118). San Francisco, CA: Jossey-Bass.
- Shadish, W. R., Jr., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: SAGE.
- Shadish, W. R., Jr., & Luellen, J. K. (2004). Donald Campbell: The accidental evaluator. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 80–87). Thousand Oaks, CA: SAGE.
- Shapiro, J. P., Secor, C., & Butchart, A. (1983). Illuminative evaluation. *Educational Evaluation and Policy Analysis*, 5(4), 465–471.
- Smith, M., & Bissell, J. (1970). Report analysis: The impact of Head Start. *Harvard Educational Review*, 40, 51–104.
- Smith, M. F. (1989). *Evaluability assessment: A practical approach*. Norwell, MI: Kluwer.
- Speller, V., Learmonth, A., & Harrison, D. (1997). The search for evidence of effective health promotion. *BMJ*, 315(7104), 361–363.
- Spencer, L. M., Schooley, M. W., Anderson, L. A., Kochtitzky, C. S., DeGroff, A. S., Devlin, H. M., & Mercer, S. L. (2013). Seeking best practices: A conceptual framework for planning and improving evidence-based practices. *Preventing Chronic Disease*, 10, 130–186.
- Spoth, R. L., & Greenberg, M. T. (2005). Toward a comprehensive strategy for effective practitioner-scientist partnerships and larger-scale community health and well-being. *American Journal of Community Psychology*, 35(3–4), 107–126.
- Stake, R. E. (Ed.). (1975). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Merrill.
- Stephenson, J., & Imrie, J. (1998). Why do we need randomised controlled trials to assess behavioural interventions? *BMJ*, 316(7131), 611–613.
- Stimson, G. V. (1996). Has the United Kingdom averted an epidemic of HIV-1 infection among drug injectors? *Addiction*, 91(8), 1085–1088.
- Strecher, V. J., Bauman, K. E., Boat, B., Fowler, M. G., Greenberg, R. A., & Stedman, H. (1989). The development and formative evaluation of a homebased intervention to reduce passive smoking by infants. *Health Education Research*, 4(2), 225–232.
- Strecher, V. J., & Rosenstock, I. (1997). The health belief model. In K. Glanz, F. M. Lewis, & B. K. Rimer (Eds.), *Health behavior and health education: Theory, research, and practice* (2nd ed., pp. 41–59). San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. L. (Ed.). (2001). *Evaluation models* (New Directions for Evaluation, No. 89). San Francisco, CA: Jossey-Bass.
- Suchman, E. (1967). *Evaluation research*. New York, NY: Russell Sage Foundation.

- Swinney, J. E., & Dobal, M. T. (2008). Embracing the challenge: Increasing workforce diversity in nursing. *Hispanic Health Care International*, 6(4), 200–204.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: SAGE.
- Tilford, S. (2000). Evidence-based health promotion. *Health Education Research*, 15(6), 659–663.
- Trochim, W. M. K. (1984). *Research design for program evaluation: The regression discontinuity approach*. Beverly Hills, CA: SAGE.
- Trochim, W. M. K. (1998). An evaluation of Michael Scriven's *Minimalist theory: The least theory that practice requires*. *American Journal of Evaluation*, 19(2), 243–249.
- Trochim, W. M. K., & Cook, J. A. (1992). Pattern matching in theory-driven evaluation: A field example from psychiatric rehabilitation. In H.-T. Chen & P. H. Rossi (Eds.), *Using theory to improve program and policy evaluations* (pp. 49–69). Westport, CT: Greenwood.
- Turner, N. C., Chen, H.-T., Harvey, D., Smith, J., Jr., & Redding, K. C. (2014). A liaison-based academic health department in Georgia: A partnership for improving community health. *Journal of Public Health Management and Practice*, 20(3), E1–E5.
- United Way of America. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: United Way of America.
- Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and Program Planning*, 45: 127–139.
- Van den Hoek, J. A., van Haastrecht, H. J., & Coutinho, R. A. (1989). Risk reduction among intravenous drug users in Amsterdam under the influence of AIDS. *American Journal of Public Health*, 79(10), 1355–1357.
- Veniegas, R. C., Kao, U. H., & Rosales, R. (2009). Adapting HIV prevention evidence-based interventions in practice settings: An interview study. *Implementation Science*, 4, 76.
- Vlahov, D., Junge, B., Brookmeyer, R., Cohn, S., Riley, E., Armenian, H., & Beilenson, P. (1997). Reductions in high-risk drug use behaviors among participants in the Baltimore needle exchange program. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 16(5), 400–406.
- Waldo, G. P., & Chiricos, T. G. (1977). Work release and recidivism: An empirical evaluation of a social policy. *Evaluation Review*, 1(1), 87–108.
- Wallerstein, N., & Bernstein, E. (1994). Introduction to community empowerment, participatory education, and health. *Health Education & Behavior*, 21(2), 141–148.
- Wallin, E., Bremberg, S., Haglund, B., & Holm, L. E. (1993). Cancer prevention in schools: Design and pilot testing of a nutritional curriculum for mid-adolescents. *Journal of Cancer Education*, 8(2), 145–150.
- Wandersman, A. (2003). Community science: Bridging the gap between science and practice with community-centered models. *American Journal of Community Psychology*, 31(3–4), 227–242.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L. . . . Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 41(3–4), 171–181.

- Wandersman, A., Valois, R., Ochs, L., de la Cruz, D. S., Adkins, E., & Goodman, R. M. (1996). Toward a social ecology of community coalitions. *American Journal of Health Promotion*, 10(4), 299–307.
- Weber, M. (1947). *The theory of social and economic organization*. (A. M. Henderson & T. Parsons, Trans.). New York, NY: The Free Press (Originally published in Germany, 1925)
- Weiner, B. J., Helfrich, C. D., Savitz, L. A., & Swiger, K. D. (2007). Adoption and implementation of strategies for diabetes management in primary care practices. *American Journal of Preventive Medicine*, 33(1 Suppl.), S35–S44; quiz S45–49.
- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell et al. (Eds.), *New approaches to evaluating community initiatives: Vol 1. Concepts, methods, and contexts*. Washington, DC: Aspen Institute.
- Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, 21(4), 501–524.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wholey, J. S. (1979.) *Evaluation: Promise and performance*. Washington, DC: Urban Institute Press.
- Wholey, J. S. (1987). Evaluability assessment: Developing program theory. In L. Bickman (Ed.), *Using program theory in evaluation* (New Directions for Program Evaluation, No. 33, pp. 77–92). San Francisco, CA: Jossey-Bass.
- Wholey, J. S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15–39). San Francisco, CA: Jossey-Bass.
- Wholey, J. S. (2004). Evaluability assessment. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 33–62). San Francisco, CA: Jossey-Bass.
- Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation* (3rd ed.). San Francisco, CA: Jossey-Bass.
- Will, K. E., & Sabo, C. S. (2010). Reinforcing Alcohol Prevention (RAP) program: A secondary school curriculum to combat underage drinking and impaired driving. *Journal of Alcohol and Drug Education*, 54(1), 14–36.
- Williams, S. G., & Redd, S. C. (2006). From research to reality: From the National Cooperative Inner-City Asthma Study to the inner-city asthma implementation. *Annals of Allergy, Asthma, & Immunology*, 97(1 Suppl. 1), S4–S5.
- Witkin, B. R., & Altschuld, J. W. (1995). *Planning and conducting needs assessments: A practical guide*. Thousand Oaks, CA: SAGE.
- Witte, K., Meyer, G., & Martell, D. (2001). *Effective health risk messages: A step-by-step guide*. Thousand Oaks, CA: SAGE.
- World Health Organization (WHO). (1998). *Health promotion evaluation: Report of the WHO European Working Group on Health Promotion Evaluation*. Copenhagen, Denmark: WHO.
- Worrall, J. L., & Gaines, L. K. (2008). The effect of police-probation partnerships on juvenile arrests. *Journal of Criminal Justice*, 34(6), 579–589.

- Wyatt Knowlton, L., & Phillips, C. C. (2013). *The logic model guidebook: Better strategies for great results* (2nd ed.). Thousand Oaks, CA: SAGE.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: SAGE.
- Zald, M. N. (1963). Comparative analysis and measurement of organizational goals: The case of correctional institutions for delinquents. *The Sociological Quarterly*, 4(3), 206–230.
- Zaltman, G., & Duncan, R. (1977). *Strategies for planned change*. New York, NY: Wiley.

Index

- Academic Health Department, 61, 62(fig.)
- Accountability
 - mature implementation stage evaluation and, 40
 - program implementation and, 180–181
 - program-monitoring systems and, 209
 - theory-driven outcome evaluation and, 305
- Accountability-oriented assessment, 7
- Accuracy, program evaluation and, 392
- Accuracy standards, 17
- Action model, 68
 - applied to anti-secondhand smoking program, 373–376
 - applied to antismoking program, 84
 - applied to community health initiative, 357, 358(fig.), 360
 - applied to diversity enhancement project, 345–348, 349(fig.), 351(fig.)
 - clarifying in program theory, 380
 - components of, 74–79
 - conceptual framework, 118, 158
 - theory-driven outcome evaluation and, 310
- Action model/change model schema, 66–70
 - advantages of using, 84–86
 - applications, 82–84, 92
 - applied to community health initiative, 355–361, 358(fig.)
 - applied to diversity enhancement project, 345–350, 349(fig.), 350(fig.), 351(fig.)
 - clarifying, 355–359
 - components of action model, 74–79
 - components of change model, 70–74
 - comprehensive form, 81(fig.)
 - conceptual framework, 70(fig.)
 - constructive assessment and, 274
 - defined, 66
 - descriptive assumptions, 66–67
 - determinants and types of, 109–110
 - exhibited generalization via, 399
 - guide to productively applying, 361–362
 - helping stakeholders gear up (or clear up), 86–91
 - prescriptive assumptions of, 67, 68–70
 - relationships among components of, 79–82
 - theory-driven process evaluation and, 191
- Action plan
 - application, 133–141
 - associate organizations/community partners, 127–128
 - clarifying and strengthening existing, 119
 - commentary or advisory approach, 146–149
 - communicating about, for future evaluation design, 119
 - conceptual framework of, 118(fig.)
 - conceptualization facilitation approach, 120, 122–133, 148
 - constructive process evaluation and modifying or clarifying, 177–179
 - defined, 98
 - developing new, 119
 - ecological context, 128–130
 - formative research approach, 120–122, 148
 - garbage reduction program, 134, 135(fig.)
 - HIV-prevention program, 138(fig.), 139–141
 - implementing organization, 123–124
 - intervention and service delivery protocols, 124–126
 - moving from program scope to, 115
 - overview, 117–120
 - pilot-testing approach, 120, 141–146, 148–149
 - program implementers, 126–127
 - program scope and, 117–119, 118(fig.)
 - questions to inform evaluator's commentary on, 147–148
 - stakeholders and, 118
 - strategies for developing, 120
 - target population, 130–133
- Action program, 3
- Action steps, 107
- Action theory, in intervening mechanism evaluation, 319, 320, 320(fig.)
- Action theory success, 382–383, 382(fig.)
- Activities, logic model, 59
- Adjuvants, 280
 - assessing joint effects of intervention and, 284–285
 - collectivity-value, 294
 - contextualizing intervention and, 286–287
 - incorporating into real-world evaluation, 268–269
 - internal validity and, 298–299
 - reason needed for real-world programs, 281–283
 - social connection, 293–294
 - types of, 282–283
- Adolescent Alcohol Prevention Trial
 - change model for, 313, 314
 - intervening mechanism evaluation of, 319
- Adoption program, identifying operational goals of, 242
- Advancing Validity in Outcome Evaluation*, 262
- Advisory approach, 48, 49
 - action plan development and, 146–149

- Advisory meeting, 42
- African-American youth, assessing needs among, 102
- After-school project for South Asian ethnic adolescents, formal specification of unintended effects, 334–335
- AIDS 101 intervention, moderating mechanism evaluation of, 325, 325(fig.)
- Aid to Families with Dependent Children, 331
 - identifying unintended outcomes of, 244
- Alcohol abuse prevention program
 - micro-level contextual support and, 77
 - moderating mechanism evaluation of, 323–324, 323(fig.), 324(fig.)
 - pre-experimental outcome evaluation of, 255
- Alcohol-abuse/safe-sex campaign, moderating mechanism evaluation of, 327
- Alcohol abuse treatment program
 - adjuvants and, 283
 - in change model, 73
 - settings for, 126
- Alcohol and drug abuse prevention program
 - intervening mechanism evaluation of, 314–315, 315(fig.)
 - stakeholder theory-based, 368, 368(fig.)
- Alcohol counseling services, health status and readiness for, 79
- Alcoholics Anonymous, 125
- American Evaluation Association, 16–17, 404
- Antibullying intervention, intervention fidelity evaluation of, 183–184
- Anti-drug abuse program
 - field study detection of unintended effects of, 336
 - theory-driven process evaluation of, 193–195, 194(table)
- Anti-secondhand smoking program
 - evaluation
 - action model, 373–376
 - evaluation findings, 377–383
 - lessons learned, 383–386
 - outcome evaluation design and change model, 376–377, 376(fig.)
 - outcome evaluation results, 378, 379(fig.)
 - process evaluation design, 377
 - process evaluation results, 377–378
 - program theory underlying, 371–373
- Antismoking interventions
 - application of action model/change model schema and, 82–84, 83(fig.)
 - constructive assessment of, 271
 - intervening mechanism evaluation of, 316–317, 317(fig.)
- Applications
 - action model/change model schema, 82–84, 92
 - action plan, 133–141
 - comprehensive evaluation typology, 51–54
 - conceptualization facilitation, 133–141
 - fundamental evaluation typology, 14
 - logic model, 61–65
 - program evaluation typologies, 14
- Arthritis self-care program, comprehensive scanning in formative evaluation of, 163–164
- Asian immigrants, family counseling program for
 - formative evaluation of, 155
- Associate organizations, 76
 - action plan development and, 127–128
 - theory-driven process evaluation of, 196
- Asthma management program, efficacy evaluations of, 252–253, 391, 392–393
- Atkins, D., 405
- Attainable goals, 232
- Attrition, as threat to internal validity, 250
- Auxiliary design, to triangulate evidence, 289–291, 295–296
- Auxiliary service organizations, 128
- Background information provision
 - strategy, 47
 - formative research approach, 120–122
- Backward reasoning, 89, 90, 91
- Bandura, A., 366
- Barriers to participation, removing, 132–133
- Beeynon, J., 160
- Berk, R. A., 318
- Bias(es)
 - eliminating potential, in holistic effectuality evaluation approach, 285–286
 - evaluators and, 16
 - self-selection, 277–278
 - using quantitative design to address, 288–289, 294–295
- Bickman, L., 65, 66
- Bicycle helmet promotion campaign, intervening mechanism evaluation of, 321
- Bilateral empowerment evaluation, 43, 50, 54, 56(fig.)
 - initial implementation evaluation and, 171–174
- Biomedical research, top-down approach and, 389
- Birth control lesson, cultural sensitivity and, 113
- Black-box evaluation, 305, 310
- Bolan, G. A., 185
- Bottom-up approach, 400–401
 - social betterment/health promotion programs and, 401–403
 - top-down *vs.*, 400–401, 401(fig.), 409–410
 - transitioning to dissemination and, 400–403
 - types of interventions for, 403
 - usefulness of, 408–410
- Boundary, program, 80
- Boundary, change model, program scope and, 98–99, 99(fig.)
- Brunk, S. E., 163–164
- Budget
 - choice of intervention strategy and, 111
 - program evaluation, 20
- Butchart, A., 159
- Campbell, D. T., 248–249, 261, 265, 398
- Campbellian validity typology, 247–248
 - basis of validity and, 262
 - distinction between internal and external validity in, 248–249

- integrative cogency model
 - and, 395
- relationships with
 - experimentation evaluation approach, 264–266
- research designs for ruling out
 - threats to internal validity, 250–258
- threats to internal validity and, 249–250
- Cancer prevention program, pilot testing, 146
- Capability, assessing implementing organization, 75–76
- Capacity building, 76
 - constructive assessment and, 270–272
 - development of program
 - monitoring/evaluation systems and, 215, 221–223, 227–228
 - evaluating in HIV prevention program, 292
 - outcome monitoring and, 208
 - program-monitoring systems and, 209–210
 - target population fidelity
 - evaluation of, 187
- Cardiovascular disease prevention program, formative research on, 121–122
- Case management services, 133
- Case studies, 408
- Causal chains, intervening
 - mechanism evaluation to assess, 319–320
- Causal mechanisms
 - assessing underlying, 291
 - mediating, 311
 - moderating, 311–312
- Causal model, anti-secondhand
 - smoking program, 376–377, 376(fig.), 378, 379(fig.)
- Causal processes, descriptive
 - assumptions and, 67
- Causal relations, program theory and, 66
- Causative theory, 67. *See also* Change model
- CDC. *See* Centers for Disease Control and Prevention (CDC)
- CDSMP. *See* Stanford Chronic Disease Self-Management Program (CDSMP)
- Centers for Disease Control and Prevention (CDC)
 - evaluation of public health and social betterment programs and, 17–18
 - HIV-prevention program, 211–228
- Change, using quantitative design to assess, 288–289, 294–295
- Change model, 67
 - applied to anti-secondhand
 - smoking program, 376–377, 376(fig.)
 - applied to antimoking
 - program, 84
 - applied to diversity enhancement
 - project, 348–350, 350(fig.), 351(fig.)
 - clarifying, 359–360, 380
 - components of, 70–74
 - intervening mechanism evaluation and, 312–313, 314(fig.)
 - program scope and, 98
 - theory-driven outcome evaluation and, 310, 311
 - See also* Action model/change model schema
- Checklist, for ranking real-world evaluations, 296–298, 297(table)
- Chelimsky, E., 15, 16
- Chen, H.-T., 262
 - bottom-up approach, 400
 - community health initiative study and, 350
 - conceptual program theory failure and, 320
 - dynamic model and, 318
 - fidelity evaluation and, 183
 - fundamental evaluation typology and, 10
 - plausibility assessment and, 237
 - program theory and, 66, 70
 - theory-driven evaluation and, 25, 305
- Child abuse cases, goals for handling, 108
- Child abuse prevention program, service delivery protocols of, 75
- Child care, as barrier to participation, 133
- Child care program, target population fidelity
 - evaluation of, 188
- Chiricos, T. G., 244
- Chronic care model, 61
- Chu, P. L., 185
- Clapp, J. D., 368
- Client-processing procedures, 75
- Client readiness, 78–79
- Clients
 - adjuvants and, 282
 - process monitoring and risk
 - behavior of, 204
 - process monitoring and
 - sociodemographic backgrounds of, 204
 - readiness of, 78–79
- Clinics, as service setting, 125–126
- Cogency
 - defined, 395
 - effectual, 395–396, 407
 - transferable, 398
 - viable, 396–398, 407–408
- Collaborating organizations, 128
- Collaborative relationships
 - action model and establishment of, 76
 - action plan development and, 127–128
- Collectivity-value adjuvant, 294
- Commentary and advisory approach, 48, 49
- Commentary approach, action plan development and, 146–149
- Commentary meeting, 42
- Communication, program scope and, 100
- Communication channels, as service
 - delivery protocol concern, 75
- Community-based organizations
 - developing program monitoring/evaluation systems and, 214
 - HIV-prevention program and, 139, 140, 306–308
- Community center, as service setting, 126
- Community coalition programs,
 - bilateral empowerment evaluation and, 171
- Community health advisors,
 - anti-secondhand smoking
 - program and, 371–372
 - lessons learned from, 383–385
 - program viability and, 381–382
 - protocol, 374
 - self-efficacy and, 380–381, 383
 - service delivery, 374–375
 - training, 377–378

- Community health initiative,
 - addressing problems in use of logic model, 350–361
 - applying action model/change model schema, 355–361, 358(fig.)
 - applying logic model, 352–355, 353(fig.)
 - project description, 350–352
- Community mobilization, action plan development and, 130
- Community organization capacity, 396n
- Community park use, constructive assessment of, 272–273
- Community partners
 - in action model, 76
 - action plan development and, 127–128
 - theory-driven process evaluation of, 196
- Community prevention alliance,
 - bilateral empowerment evaluation of, 172–173
- Competence, evaluator, 17
- Comprehensive evaluation
 - typology
 - action model/change model schema and, 92
 - application of, 51–54
 - development strategies, 47–50
 - dynamics of transition across stages, 41
 - evaluation approaches associated with each stage, 41–44
 - initial implementation stage, 37(table), 40, 42–43
 - mature implementation stage, 38(table), 40, 43
 - merit assessment strategies, 46–47
 - multiple-entry evaluation, 54–55, 56–57
 - outcome stage, 38(table), 40–41, 43–44
 - program-planning stage, 37(table), 39–40, 42
 - single-entry evaluation, 55–56
 - stages in program life cycle, 36–41, 37–38(table)
 - strategies underlying evaluation approaches, 44–50
- Comprehensive scanning,
 - in formative evaluation, 162–164
- Compulsory seat-belt legislation
 - effect, formal specification of possible unintended effects, 334
- Concept mapping
 - example, 178–179
 - mature implementation evaluation and, 178
- Conceptualization facilitation, 49, 103–112
 - action plan development, 122–133, 148
 - advantages of, 136, 140–141
 - application, 133–141
 - associate organizations/ community partners, 127–128
 - choosing interventions/treatments that affect the determinant, 110–112
 - clarifying program theory using, 309–310
 - determinants and types of action model/change model schemas, 109–110
 - ecological context, 128–130
 - implementing organization, 123–124
 - intervention and service delivery protocols, 124–126
 - mature implementation evaluation and, 178
 - program implementers, 126–127
 - target population, 130–133
 - theorizing methods, 104–109
 - working group or intensive interview format, 103–104
- Conceptual theory, in intervening mechanism evaluation, 320, 320(fig.)
- Conceptual theory success, 382–383, 382(fig.)
- Conclusion-oriented evaluation, 82
- Conclusive assessment, 23
- Conclusive/constructive outcome evaluation, 10(fig.), 14
- Conclusive/constructive process evaluation, 10(fig.), 13
- Conclusive outcome assessment, 269–270, 275–280
 - of HIV prevention program, 293–294
 - rating system, 296–298, 297(table)
- Conclusive outcome evaluation, 10(fig.), 12–13, 44, 245
 - outcome monitoring, 44
 - purpose of, 230
 - validity-focused outcome evaluation, 44
 - viability evaluation, 44
- Conclusive process evaluation, 10(fig.), 11–12, 43
 - conceptual framework, 180(fig.)
 - designing, 181–182
 - fidelity *vs.* “reinvention” in, 189–190
 - intervention fidelity evaluation, 183–184
 - of mature implementation, 43, 151, 180–182
 - process monitoring, 43
 - referral fidelity evaluation, 184–185
 - selecting evaluation approaches, 182
 - service delivery fidelity evaluation, 185–186
 - target population fidelity evaluation, 186–189
- Concurrent cogency approaches, 406–408
- Condom distribution program,
 - moderating mechanism evaluation of, 325–326, 326(fig.)
- Conflicts of interest, multiple-entry evaluation and, 54–55
- Conflict Tactics Scale, 257
- Consensus-building approach,
 - plausibility assessment, 231, 237–245
- Consortium, action plan
 - development and joining/ launching, 129
- Construction-oriented evaluation, 82
- Constructive evaluation
 - bilateral empowerment evaluation, 43
 - formative evaluation, 42, 43
 - formative research, 42
 - of initial implementation, 42–43
 - for initial implementation evaluation, 151
 - logic models and program theory, 42
 - of mature implementation, 43
 - needs assessment, 42

- for planning stage, 42
- program review/development meeting, 43
- Constructive outcome assessment, 269–275
 - of HIV prevention program, 292–293
 - participatory evaluation, 272–274
 - rating system, 296–298, 97(table)
 - using program theory or logic models to check for vulnerabilities, 274–275
- Constructive outcome evaluation, 10(fig.), 12, 43–44, 230–245
 - evaluability assessment, 43, 231, 234–236
 - overview, 230–231
 - plausibility assessment/consensus-building approach, 44, 231, 237–245
 - purpose of, 230
 - SMART goals, 43, 231–234
- Constructive process evaluation, 10(fig.), 11
 - bilateral empowerment evaluation, 171–174
 - formative evaluation approach, 154–165
 - for initial implementation evaluation, 153–174
 - for mature implementation evaluation, 176–180
 - modifying or clarifying program scope or action plan, 177–179
 - program review/development meeting, 165–171
 - troubleshooting implementation problems, 179–180
- Construct validity, 248
 - theory-driven outcome evaluation and, 305
- Consumer Reports*, 337
- Contextualizing intervention, 293
 - in real-world setting, 286–288
- Contextual support, 77–78
- Contingency view on methods,
 - bottom-up approach and, 410
- Conybeare, J. A., 334
- Cook, T. D., 248, 289
- Coordinated Approach to Child Health, effectiveness evaluation of, 390
- Core organizations, 128
- Cost
 - efficacy evaluation, 409
 - evaluation, tradeoff with thoroughness and, 53
 - internal *vs.* external evaluation and, 15
 - outcome monitoring, 207, 208
 - program review/development meeting and, 168
 - viability evaluation, 409
- Counseling, Testing, Referral, and Partner Notification, referral fidelity evaluation of, 185
- Coverage, fidelity evaluation and, 183
- Credible evidence, 26, 403–406
 - bottom-up approach and, 409–410
 - gathering, 18
 - randomized controlled trials and, 301
- Crime prevention program, logic model, 63–65, 64(fig.)
- Cronbach, L. J., 8, 15, 23, 50, 261, 262
- Cultural appropriateness, 396n
 - adjuvants and, 283
- Cultural change, 347
- Cultural competence, choosing implementing organization and, 123–124
- Cultural sensitivity, target population and, 113
- Culture, as barrier to participation, 132
- Curriculum evaluation, action model/change model schema and, 85
- Dagenais, C., 178
- DARE program, 394, 396
- Data aggregation, at national level, 218–219
- Data collection
 - comprehensive scanning and, 162
 - constructive outcome assessment and, 272
 - creating action plan and, 121
 - formative evaluation and, 156–157
 - integrative process/outcome evaluation and, 329
 - intervening mechanism evaluation and, 318
- pilot testing and, 143
- process monitoring *vs.* process evaluation, 205–206
- theory-driven outcome evaluation and, 310
- theory-driven process evaluation and, 191
- Delphi method, 178
- Demonstration project, 142
- Demonstration tool, modifying, 294
- Descriptive assumptions
 - action model/change model schema and, 66–67
 - program theory and, 66
- Descriptive theory. *See* Change model
- Desirability, plausibility *vs.*, 105–106
- Determinants
 - in change model, 71–72
 - choosing interventions/treatments that affect, 110–112
 - identifying, that are likely to change outcomes, 108–109
 - moderating mechanism and relationship between outcome or intervention and, 325–326, 325(fig.), 326(figs.)
 - program review/development meeting and, 170–171
 - types of action model/change model schemas and, 109–110
- Developmental evaluation approach, 31
- Development facilitation, 49, 177
 - conceptualization facilitation approach (*See* Conceptualization facilitation)
 - participatory modes for, 88–89
 - theorizing procedures for, 89–91
- Development partnership strategy, 171
 - for bilateral empowerment evaluation, 171–174
 - for initial implementation evaluation, 153–154
- Development strategies, 47–50
 - background information provision strategy, 47
 - development facilitation strategy, 49
 - enlightenment strategy, 49–50
 - partnership strategy, 50
 - troubleshooting strategy, 47–49

- Diagnostic assessment, 131
 Diaz, R. M., 306–308
 Diffusion tradition, 189
 Dimension test, 400
 Direct observation, 288
 Discussion, rough draft to facilitate, 91
 Dissemination
 bottom-up approach to, 400–403
 top-down approach to, 388–389
 Diversity enhancement project, addressing problems in use of logic model, 344–350
 applying action model/change model schema, 345–350, 349(fig.), 350(fig.), 351(fig.)
 applying logic model, 345, 346(fig.)
 project description, 344
 Division of labor in service delivery, 75
 Documents/literature, service delivery and, 125
 Donaldson, S. L., 262, 313, 324–325
 Doran, G. T., 232
 Drop-out rates, 131–132
 Drug abuse prevention/intervention programs
 defined outcomes for, 107
 establishing goal for, 106
 intervening mechanism evaluation of, 314–315, 315(fig.)
 stakeholder theory-based, 368, 368(fig.)
 Durkheim, Émile, 69
 Dutton, D. G., 257
 Dynamic model of intervening mechanism evaluation, 317–318, 318(fig.)
 Early, T. J., 368
 Ecological context
 action model and, 77–78
 action plan development and, 128–130
 for anti-secondhand smoking program, 375–376
 pilot testing and, 145–146
 program review/development meeting and, 170
 Ecological intervention program, 78
 Economic self-sufficiency for low-income women program, program review/development meeting in initial implementation evaluation, 169–171
 Educational program
 action plan development and, 129
 as dynamic model, 317, 318(fig.)
 Education/nonsmoking
 signage intervention, 371, 372–373
 evaluation findings, 377–383
 lessons learned, 383–385
 outcome evaluation design and change model, 376–377, 376(fig.)
 outcome evaluation results, 378, 379(fig.)
 process evaluation results, 378
 protocol, 373–374
 service delivery, 374
 Educators for Social Responsibility, 397
 Effectiveness evaluation, 389, 400.
 See Real-world outcome evaluation
 Effectual cogency, 395–396
 focusing on, 407
 Effectuality
 in integrative cogency model, 395
 outcome evaluation and, 381
 transferability and, 405
 Effectuality evaluation, 398
 Efficacy
 anti-secondhand smoking program and, 380–381, 384–385
 formal theory-based interventions and, 369
 transferability and, 405
 Efficacy evaluation, 252–253, 400
 effectiveness outcome evaluation *vs.*, 263–264
 internal validity and, 391
 top-down approach and, 389, 390–391
 See also Randomized controlled trials (RCTs)
 Efficacy expectation, 366, 367(fig.)
 Elderly service program, constructive outcome evaluation of, 231
 Electronic data system, 209
 Eligibility criteria
 action model and, 78–79
 action plan development and, 130
 constructive outcome assessment of, 271
 Enlightenment-oriented evaluation, 82
 Enlightenment strategy, 49–50
 Environment, program, 4–5. *See also* Ecological context
 Environmental change or trends, adjuvants and taking advantage of, 283
 Environmental Protection Agency (EPA), 13
 Environmental scan, 287
 EPA. *See* Environmental Protection Agency (EPA)
 Eriksen, M. P., 192
 Ethics, program evaluation, 16–17
 Ethridge, M. E., 331
 Evaluability assessment, 43, 231, 234–236
 Evaluation practitioners, use of book, 32
 Evaluators
 internal *vs.* external, 14–15
 role in bilateral empowerment evaluation, 173–174
 Evidence
 building from ground up, 301–302
 choice of intervention strategy and, 111–112
 efficacious, *vs.* real-world effectiveness, 380–381
 in integrative cogency model, 395
 types of, to support evidence-based practice, 395
 using auxiliary design to triangulate, 289–291, 295–296
 See also Credible evidence
 Evidence-based intervention movement, 370–371
 increase in popularity of, 403–404
 Evidence-based interventions/evidence-based practice dissemination of, 390
 limitations and strategies to address them, 403–406
 modification to fit specific situation, 393–394
 real-world operations and, 390–394

- Evidence gap, 301–302
- Exhibited generalization, 399
- Expectancy effects, randomized controlled trials and, 261
- Experimental designs, for outcome evaluation, 251–253
 - case illustration of, 252–253
- Experimentation evaluation
 - approach, 247–258
 - conceptualization of outcome evaluation, 267
 - efficacy evaluation and, 263
 - experimental designs, 251–253
 - foundation of, 247–248
 - holistic effectuality evaluation approach *vs.*, 299–301
 - internal validity *vs.* external validity, 248–249
 - ongoing debates over, 260–262
 - pre-experimental designs, 253–255
 - quasi-experimental designs, 256–258
 - relationships with Campbellian validity typology, 264–266
 - research designs for ruling out threats to internal validity, 250–258
 - threats to internal validity, 249–250
- External evaluators, internal evaluators *vs.*, 14–15
- External feedback loops, in action model/change model schema, 82
- External purpose of stakeholders, matching evaluation strategy or approach to, 51–53
- External validity, 261, 262
 - in Campbellian validity typology, 248–249
 - transferable cogency and, 398
- Family counseling program
 - formative evaluation of, 155
 - intervention protocol of, 75
 - moderating mechanism evaluation of, 321–322
- Feasibility
 - program evaluation and, 392
 - standards, 17
- Feasibility study, 141. *See also* Pilot testing
- Feedback
 - program, 5–6
 - stakeholders and, 5, 7
- Feedback loops, in action model/change model schema, 81–82
- Fetterman, D. M., 85
- Fidelity evaluation, 46, 182–190
 - intervention, 183–184
 - referral, 184–185
 - “reinvention” *vs.*, 189–190
 - service delivery, 185–186
 - target population, 186–189
- Field study detection of unintended effects, 335–336
- Field-testing, 112
- Flay, B. R., 263
- Fluid complexity, 30–31
- Flu immunization program,
 - constructive outcome assessment of, 271
- Focus group
 - formative evaluation and, 48, 155, 157, 160–161
 - formative research and, 102, 408
 - holistic effectuality evaluation and, 287
- Food policy coalition, intensive interviews in formative evaluation of, 161–162
- Food relief programs, 73
- Formal specification of possible unintended effects, 334–335
- Formal theory
 - as basis for intervention programs, 365–371, 367(fig.), 368(fig.)
 - case study, 371–379, 376(fig.), 379(fig.), 383–386
 - integrating with stakeholder theory, 386
 - strengths and limitations of, 379–383, 382(fig.)
- Formal theory-based interventions, 365–366
 - conceptual theory success and, 382–383
 - relative value of, 369–371
 - stakeholder theory-based interventions *vs.*, 365–368
 - viability issues, 386
- Formative evaluation, 42, 43, 48
 - comprehensive scanning, 162–164
 - defined, 8
 - focus group meeting, 160–161
 - for initial implementation evaluation, 154–165
 - intensive interviews, 161–162
 - on-site observation and checking, 158–160
 - pilot testing and, 154–155
 - research methods, 155–156
 - steps in applying, 156–158
 - summative evaluation *vs.*, 7–9
 - timeliness and relevancy, 155
 - types of, 158–164
 - use of results, 164–165
- Formative research, 42, 47, 48, 154
 - action plan and, 120–122, 149
 - program scope and, 102–103
- Fort Bragg Child and Adolescent Mental Health Demonstration, integrative process/outcome evaluation of, 329–331, 330(fig.)
- Forward/backward reasoning, 91
- Forward reasoning, 90
- Framework of Program Evaluation for Public Health*, 17–18
- Freeman, J. D., 169, 170
- Funding agencies
 - logic models and, 343
 - process-monitoring data and, 205
- Garbage reduction program, action plan development, 133–136, 135(fig.)
- Garbe, P., 262
- Gargani, J., 262
- Generalization
 - exhibited, 399
 - targeted, 400
- Gettleman, L., 121
- Glasgow, R. E., 188, 263
- Goal attainment, 396
- Goal-free evaluation, 238
- Goals
 - in change model, 71
 - constructive outcome evaluation and development of, 231
 - disagreement about which goals to evaluate, 238–239
 - identifying final, 105–108
 - identifying in outcome monitoring, 206
 - objectives *vs.*, 107
 - outcomes *vs.*, 107
 - short-term *vs.* long-term, 106–107
 - SMART, 231–234, 268, 270

- ultimate, 106
- See also* Plausibility assessment/consensus-building
- Goal trap, 237–238
- Goeppinger, J., 163–164
- Goodman, R. M., 172–174
- Gottlieb, N. H., 192, 335
- Government Performance and Results Act (1993), 209
- Gowdy, E. A., 169, 170
- Gowen, L. K., 306–308
- Graham, J. W., 313
- Grant applications, logic model and, 343
- Green, L. W., 192
- Greene, J., 262
- Guba, E. G., 22, 49
- Hansen, M. B., 313
- Hawe, P., 161
- Haynes, B., 405
- Head Start, 24
 - disagreement on which goals to evaluate, 239
 - eligibility criteria, 78
- Health belief model, 306
- Health education program for pregnant women, troubleshooting implementation problems in, 179–180
- Health promotion programs
 - adjuvants and, 281
 - bottom-up approach and, 401–403
 - credible evidence of, 405
 - determinants and, 71–72
 - distinction between operational and official goals and, 241
 - evidence-based intervention movement and, 404
 - holistic effectuality evaluation of, 290
 - integrated evaluation of, 27
 - top-down approach and, 389
- Health-related programs, treatment in, 73–74
- Health status, client readiness and, 78–79
- Heart disease and stroke reduction program, logic model, 61, 63(fig.)
- Heath belief model, 72
- Hirozawa, A. M., 185
- History threat to internal validity, 249, 279–280, 282–283
- HIV counseling and testing
 - intervention, clarifying stakeholder intentions in, 182
- HIV-positive population, outcome monitoring of program to reduce intolerance toward, 206
- HIV-positive speaker presentation, implicit theory for, 306–307, 307(fig.)
- HIV-prevention programs
 - action plan development and, 130, 136–141, 138(fig.)
 - adaptation of evidence-based intervention and, 393
 - barriers to development of national monitoring/evaluation system, 214–215
 - building momentum for collaboration, 215–218
 - change in mature implementation stage and, 189–190
 - constructive assessment of, 273
 - developing program monitoring/evaluation system for, 211–228
 - eligibility criteria, 78
 - evaluation guidance, 218–221
 - evidence-based interventions and, 384
 - formative evaluation of, 48
 - health status and readiness for, 79
 - intervention and service delivery protocols for, 124–125
 - making stakeholder theory explicit in, 306–308, 307(fig.)
 - mediating causal mechanisms of, 311
 - moderating causal mechanisms of, 312
 - outcome of, 71
 - real-world outcome evaluation of, 291–296
 - relationships among components of action model and, 80
 - stakeholder participation, 213–214
 - stakeholder participation issues, 223–226
 - technical assistance and capacity building in, 221–223, 227–228
 - theory-driven process evaluation of mature implementation of, 195–198
 - threats to internal validity of evaluation of, 249
- HIV-testing/counseling program, referral fidelity evaluation of, 185
- HIV transmission, identifying target population, 105
- Hofmann, D. A., 289
- Holistic assessment, 21
 - action model/change model schema and, 84–85
- Holistic effectuality evaluation
 - approach, 266–269
 - addressing issues through hybrid evaluation, 267–268
 - adjuvants and, 281–283
 - building evidence from ground up, 301–302
 - checklist for ranking real-world evaluations, 296–298, 297(table)
 - conceptualization of outcome evaluation, 267–269
 - conclusive assessment, 269–270, 275–280
 - constructive assessment, 269–275
 - experimentation evaluation approach *vs.*, 299–301
 - incorporating adjuvants into, 268–269
 - methodology, 284–286
 - research steps for assessing real-world effects, 286–296
 - usefulness of, 298–299
- Hollis, J., 188
- Homeless program, goals for, 106
- Hospitals, as service setting, 125–126
- House, E. R., 16, 262
- Housing discrimination prevention program, differentiating plausible goals from window-dressing goals, 240
- Hybrid outcome evaluation, 44, 267–268
 - conclusive/constructive, 10(fig.), 14
- Hybrid process evaluation, 43, 44, 151, 190–198
 - conclusive/constructive, 10(fig.), 13

- Hybrid types of program evaluation, 10(fig.), 13–14, 24, 44
 - commentary or advisory meeting, 42
 - enlightenment strategy and, 50
 - pilot testing, 42
 - for planning stage, 42
 - relevancy testing, 42
 - theory-driven process evaluation, 43, 44, 151, 190–198
 - theory-drive outcome evaluation, 44
 - transferability evaluation, 44
- If . . . then statements, 59
- Implementation, action model/
 - change model schema and, 92. *See also* Initial implementation; Mature implementation
- Implementation evaluation, 11, 13, 151. *See also* Initial implementation evaluation; Mature implementation evaluation
- Implementation stage, 19
- Implementers. *See* Program implementers
- Implementing organization
 - action plan development and capacity of, 123–124
 - for anti-secondhand smoking program, 375
 - assessing capabilities of, 75–76
 - pilot testing and, 145
 - program review/development meeting and, 169
 - theory-driven process evaluation of, 195
- Implicit theory
 - clarifying, 306–309, 307(fig.), 308(fig.)
 - stakeholder theory and, 367
- Improvement-oriented assessments, 7
- Incentives, for participation in programs, 132
- Independence, internal *vs.* external evaluation and, 14–15
- In-depth interviews, in holistic effectuality evaluation approach, 287
- Indigenous evaluation theories and methodologies, 29, 299
- Initial implementation evaluation,
 - constructive process evaluation, 153–174
 - bilateral empowerment evaluation, 171–174
 - formative evaluation approach, 154–165
 - program review/development meeting, 165–171
- Initial implementation stage, 19
 - in comprehensive evaluation typology, 37(table), 40, 42–43
- Inner-City Asthma Interventions, efficacy evaluation of, 393
- Innovation, diffusion of, 396n
- Inputs
 - logic model, 59
 - program, 4
- Instrumentation threat to internal validity, 250, 278–279
- Integrated evaluation perspective, 26–29, 394–395
 - bottom-up approach and, 400–403
 - on concurrent cogency approaches, 406–408
 - See also* Integrative cogency model
- Integrative cogency model, 394–398
 - credible evidence and, 403–406
 - effectual cogency, 395–396
 - effectuality evaluation, 398
 - focusing on effectual cogency, 407
 - focusing on viable cogency, 407–408
 - optimizing approach, 408
 - transferability evaluation, 399–400
 - transferable cogency, 398
 - usefulness of, 408–410
 - viability evaluation, 399
 - viable cogency, 396–398
- Integrative process/outcome evaluation, 327–333
 - examples of, 329–333, 330(fig.), 332(fig.)
 - linkages of major components in, 328, 328(fig.)
 - research methods and strategies associated with, 329
- Integrative viability model, 386
- Integrity in evaluation process, 17
- Intensity, fidelity evaluation and, 183
- Intensive interviews, 88–89, 408
 - conceptualization facilitation and, 103–104, 122
 - formative evaluation and, 161–162
- Internal evaluators, external evaluators *vs.*, 14–15
- Internal feedback loops, in action model/change model schema, 81–82
- Internal purpose of stakeholders, matching evaluation strategy or approach to, 51–53
- Internal-use evaluation, 82
- Internal validity
 - in Campbellian validity typology, 248–249
 - credible evidence debates and, 404–405
 - effectiveness evaluation and, 263–264
 - efficacy evaluation and, 390–391, 391
 - history threat to, 279–280, 282–283
 - instrumentation threat to, 278–279
 - maturation threat to, 279, 283
 - real-world evaluation and elimination of threats to, 285
 - relationship between adjuvants and, 298–299
 - research designs for ruling out threats to, 250–258
 - selection threat to, 283
 - self-selection bias and, 277–278
 - testing threat to, 282
 - theory-driven outcome evaluation and, 305
 - threats to, 249–250, 277–280
 - as top priority in evaluations, 261
 - types of adjuvants and threats to, 282–283
- Interrupted time-series design, 257–258, 289
- Intervening mechanism evaluation, 312–321
 - basic model, 312(fig.)
 - dynamic model, 317–318, 318(fig.)
 - linear model, 313–314
 - multiple-determinant model, no sequential order, 315, 316(fig.)

- multiple-determinant model,
 - sequential order, 315–316, 317(fig.)
- one-determinant model, 314–315, 315(fig.)
- theoretical bases of, 318–319
- when to use, 319–321, 320(fig.)
- Intervening variable, 71
- Intervention
 - adjuvants and enhancing, 282
 - assessing joint effects of adjuvants and, 284–285
 - in change model, 73–74
 - choosing, that affect determinant, 110–112
 - components, assessing
 - effectiveness of, 320–321
 - contextualizing, 286–288, 293
 - defined, 347
 - implicit theory for, 306–308, 307(fig.)
 - moderating mechanism and relationship between determinant or outcome and, 325–326, 325(fig.), 326(figs.)
 - partnership and, 356
- Intervention activities, process monitoring and, 204
- Intervention classroom, 125
- Intervention fidelity evaluation, 183–184
 - example, 184
- Intervention programs, 3
 - population-based, 3*See also* Program
- Intervention protocols
 - action model and, 74–75
 - action plan development and, 124–126
 - anti-secondhand smoking program, 373–374
 - in pilot testing, 144
- Interviews, 102
 - clarifying stakeholders' implicit theory and, 308
 - in-depth, 287*See also* Intensive interviews
- Ituarte, P., 321
- Job-training program
 - action model and, 74
 - eligibility criteria, 78
- Joint Committee on Standards for Education, 17
- Julnes, G., 262
- Junkiebond (Junkie League), 402
- Juvenile crime reduction/prevention programs
 - determinants and, 72, 104
 - identifying operational goals of, 242
 - interrupted time-series evaluation of, 258
 - intervention/treatment and, 73
 - selection of determinants and interventions, 108–109
- Kaftarian, S. J., 85
- Katz, M. H., 185
- Key informant interviews, in
 - formative evaluation, 155, 156
- “Key to Health,” pilot testing, 146
- Komington, 110
- Kristal, A. R., 315, 321
- La Chance, P. A., 188
- Lando, H., 188
- Language, as barrier to participation, 132
- Large-scale programs, disagreement on which goals to evaluate, 238–239
- Laub, C., 306–308
- Leadership- and management-training program, on-site observation and checking of, 159–160
- Learnfare, integrative process/ outcome evaluation of, 331–333, 332(fig.)
- Lenihan, K. J., 318
- Liaison-based academic health department project, logic model, 61, 62(fig.)
- Lincoln, Y. S., 22, 49
- Linear model of intervening mechanism evaluation, 313–314
- Logic models, 19, 42, 58–65
 - action to take in event of failure of, 343–363
 - application examples, 61–65
 - applied to planning evaluation, 352–355
 - checking for potential vulnerabilities using, 274–275
 - community health initiative, 350–361, 353(fig.), 358(fig.)
 - diversity enhancement project, 344–350, 346(fig.), 349(fig.), 350(fig.), 351(fig.)
 - productively applying logic model and action model/change model schema, 361–362
 - systems change and evaluation in the future, 362–363
 - United Way's, 59, 60(fig.)
- Logistical barriers, 132–133
- Long-term goals, 106–107
- Lovato, C. Y., 192
- Low-income women
 - initial implementation evaluation of economic self-sufficiency program for, 169–171
 - targeting for cardiovascular disease prevention program, 121–122
- Low-interest loans for farmers
 - program, action model/change model schema and, 69, 85
- Macrolevel contextual support, 77
 - action plan development and, 129
- Management support, for program-monitoring systems, 210, 227
- Mandatory programs, target populations for, 132
- Manpower, choosing implementing organization and, 124
- Mark, M. M., 262, 289
- Marx, R., 185
- Mass media, service delivery and, 125
- Maturation threat to internal validity, 249, 279, 283
- Mature implementation evaluation, 176–198
 - conclusive process evaluation for, 177, 180–182, 180(fig.)
 - constructive process evaluation for, 176–180
 - designing to fit stakeholders' needs, 181–182
 - fidelity *vs.* “reinvention” in, 189–190
 - hybrid process evaluation for, 177, 190–198
 - intervention fidelity evaluation, 182–184
 - referral fidelity evaluation, 184–185
 - selecting, 182
 - service delivery fidelity evaluation, 185–186

- target population fidelity
 - evaluation, 186–189
- theory-driven process evaluation, 190–198
- Mature implementation stage, 19
 - in comprehensive evaluation typology, 38(table), 40, 43
- McKillip, J., 290
- McRae, S. G., 188
- Measurable goals, 232
- Media campaigns
 - action plan development and, 129
 - recruiting target population and, 131
- Mediating causal mechanism, 311
- Mediating variable, 71
- Medicine, evidence-based, 404
- Mental health treatment programs, settings for, 126
- Mental readiness of client, 78
- Mercier, C., 178
- Merit assessment, 24
 - hybrid process evaluation and, 190
- Merit assessment strategies, 44, 46–47
 - performance assessment strategy, 46
 - performance-monitoring strategy, 46–47
- Mesological intervention, replication of, 291
- Method-driven evaluation, 25
- Methods
 - action model/change model schema, selection of most suitable and, 86
 - contingency view on, 410
- Microlevel contextual support, 77
 - action plan development and, 129
- Middle-school dropout program, service delivery fidelity
 - evaluation of, 186
- Moderating causal mechanism, 311–312
- Moderating mechanism evaluation, 321–327
 - advanced, 324–326, 325(fig.), 326(figs.)
 - basic model of, 322(fig.)
 - constructing, 322
 - examples, 323–324, 323(fig.), 324(fig.)
 - when to use, 327
- Moderators, 321–322
- Modes of delivery, 125
- Monitoring
 - guidance on, 209
 - logic model and, 59–60
- Mpower, 384, 393
- Multi-attribute utility method, 178
- Multilevel intervention program, 78
- Multiple-determinant model
 - no sequential order, 315, 316(fig.)
 - sequential order, 315–316, 317(fig.)
- Multiple-entry evaluation, 54, 56–57, 56(fig.)
- National Alliance of State and Territorial AIDS Directors, 216, 226
- National Cooperative Inner-City Asthma Study, efficacy
 - evaluation of, 392–393
- National Institutes of Health, 252
- Needle exchange programs, as
 - example of bottom-up approach, 402–403
- Needs assessment, 42, 47, 355
 - clarifying role of, 356
 - program scope and, 101–102
- Neighborhood Watch group, 63–65, 64(fig.)
- New York City Board of Education, 397–398
- Nominal group technique, 178
- Nonequivalent comparison group
 - design, 256–257, 288–289
- Normative theory, 68
- “The Not AEA Statement,” 404
- Nursing care program, eligibility
 - criteria, 78
- Nursing student diversity program. *See* Diversity enhancement project
- Nutrition program, intervening
 - mechanism evaluation of, 321
- Objectives, goals *vs.*, 107
- Observation
 - direct, 288
 - participant, 288
- Observational methods, in holistic
 - effectuality evaluation approach, 288
- Occupational Safety and Health Administration (OSHA), 13
- Office/clinic/hospital, as service
 - setting, 125–126
- Official goals. *See* Plausibility assessment/consensus-building
- Ohio District Eleven Adoption Project, 242
- Olsen, L., 242
- One-determinant model of
 - intervening mechanism evaluation, 315(fig.), 3140–315
- One-group posttest-only design, 254
- One-group pretest-posttest design, 254–255, 288
 - combined with pattern matching, 289
 - expansion of, 257–258
- One-on-one interaction, 125
- On-site observation and checking, 158–160
 - example, 159–160
- Open Airways for School program, efficacy
 - evaluation of, 252–253, 391
- Operative goals, 241–243
 - defined, 241
 - methods for identifying, 242–243
- Organizational mission and
 - philosophy, choice of intervention strategy and, 111
- Organizations. *See* Associate organizations; Implementing organization
- OSHA. *See* Occupational Safety and Health Administration (OSHA)
- Osteoporosis awareness program, focus group meeting on, 160–161
- Outcome, 19
 - action model/change model schema and, 92
 - in change model, 71
 - goals *vs.*, 107
 - identifying determinants likely to change, 108–109
 - identifying measurable, 105–108
 - in logic model, 59
 - moderating mechanism and
 - relationship between determinant or intervention and, 325–326, 325(fig.), 326(figs.)
- Outcome evaluation, 46, 181
 - anti-secondhand smoking program, 378, 379(fig.)

- design for anti-secondhand smoking program, 376–377
- efficacy *vs.* effectiveness in, 263–264
- evaluability assessment and, 234
- experimentation evaluation approach, 247–258
- outcome monitoring *vs.*, 207–208
- viability and, 381–382
- See also* Conclusive outcome evaluation; Constructive outcome evaluation; Constructive process evaluation; Real-world outcome evaluation; Theory-driven outcome evaluation
- Outcome expectation, 366, 367(fig.)
- Outcome monitoring, 44, 46–47, 206–208
 - determining required *vs.* optional, 220
 - identification of goals, 206
 - outcome evaluation *vs.*, 207–208
 - outcome measures and data collection and, 207
 - process monitoring *vs.*, 207
- Outcome stage, in comprehensive evaluation typology, 38(table), 40–41, 43–44
- Output
 - logic model, 59
 - program, 4
- Outreach debriefings, 377
- Outreach workers, 131
 - training, 377–378
- Overcoverage, 130–131
- Parenting program
 - nonequivalent comparison group evaluation of, 256–257
 - one-group posttest-only outcome evaluation design, 254
 - static group comparison of, 254
- Parenting skills interventions,
 - one-group pretest-posttest outcome evaluation of, 255
- Participant observation, 48, 155, 156, 288
- Participation
 - adjuvants and encouraging, 283
 - removing barriers to, 132–133
- Participatory approach, to design of outcome evaluation, 292
- Participatory evaluation, 16, 272–274
- Participatory modes, for development facilitation, 88–89
- Participatory strategy, to develop program-monitoring system, 210–211
- Partnership, intervention and, 356
- Partnership strategy, 50
- Passive smoking by infants
 - intervention program formal theory-based, 366, 367(fig.)
 - relevancy testing of, 113–115, 114(fig.)
- Path analysis, 319
- Pattern matching
 - in evaluation of HIV-prevention program, 295–296
 - in holistic effectuality evaluation approach, 289–290, 295–296
- Patton, M. Q., 49, 85
- Pawson, R., 305
- Peer resource manual, 222
- Peladeau, N., 178
- Percy, S. L., 331
- Performance assessment, 46, 182
- Performance-monitoring strategy, 46–47
- Perinatal care program, relevancy testing approach and, 112
- Permanent evaluation system, 163
- Perrow, C., 241
- Personnel restraints, choice of intervention strategy and, 111
- Persuasion, evaluation and, 262
- Physical readiness of clients, 78–79
- Piat, M., 178
- Pilot study, 142
- Pilot testing, 42, 48–49
 - action plan development and, 141–146, 148–149
 - conducting, 142–143
 - defining, 141–142
 - designing, 143–146
 - evaluation guidance, 219
 - formative evaluation and, 154–155
- Planning stage, 19
- Plausibility, desirability *vs.*, 105–108
- Plausibility assessment/consensus-building, 44, 231, 237–245
 - discussing findings with stakeholder groups, 244–245
- identifying operative goals that do not appear in official goals, 241–243
- identifying possible unintended outcomes, 243–244
- identifying which official goals are plausible and which are not, 239–241
- phases of, 239–245
- potential problems based on official goals, 237–239
- revising and finalizing program goals and outcomes for evaluation, 245
- Politics, program evaluation and, 15–16
- Population-based intervention program, 3
- PRECEDE-PROCEED model, 72
- Pre-experimental designs, for outcome evaluation, 253–255
 - case illustration of, 255
 - one-group posttest-only design, 254
 - one-group pretest-posttest design, 254–255
 - static group comparison, 254
- Prescriptive assumptions
 - in action model/change model schema, 67, 68–70
 - program theory and, 66
- Prescriptive theory, 68
- Prisoner rehabilitation, distinction between operational and official goals and, 241, 242
- Private home, as service setting, 126
- Problems
 - identification of, 104–105, 157
 - remedial action for, 157–158
- Process evaluation, 43
 - design for anti-secondhand smoking program, 377
 - process monitoring *vs.*, 205–206
 - See also* Conclusive process evaluation; Constructive process evaluation; Implementation evaluation
- Process monitoring, 43, 46–47, 204–206
 - outcome monitoring *vs.*, 207
 - process evaluation *vs.*, 205–206
 - uses of data, 205
- Product evaluation, 21
- Program
 - boundaries, 80

- clarifying change model and action model in, 380
- clarifying intentions of, 181–182
- components of, 4–6, 58
- descriptions of (*See* Action model/change model schema; Logic models)
- environment of, 4–5
- evaluation design and background information on, 18
- feedback, 5–6
- function of, 5
- inputs, 4
- nature/characteristics of, 3–6, 4(fig.)
- outputs, 4
- as social system, 27
- transformation, 4
- Program coherence, constructive outcome evaluation to ensure, 270
- Program development
 - pilot-test findings and, 143
 - stages of, 19
- Program document review, in formative evaluation, 156, 157
- Program effectiveness, program theory and, 65
- Program evaluation
 - accountability standards, 17
 - addressing both scientific and stakeholder credibility, 22–23
 - art of, 35–36, 51
 - bottom-up approach, 400–403, 408–410
 - challenges of, 21–24
 - concepts, theories, and methodologies, 6–7
 - definition of, 5
 - determining required *vs.* optional, 220
 - developing HIV-prevention program, 211–228
 - ethics and, 16–17
 - focusing, 18
 - four standards of, 392
 - future of, 362–363
 - holistic approach to, 21
 - integrated evaluation perspective, 26–29, 406–408
 - integrative cogency model, 394–400, 403–406, 408–410
 - internal *vs.* external evaluators, 14–15
 - judging program by results and context, 21
 - major challenges of, 20–24
 - politics and, 15–16
 - program complexity and evaluation theories, 30–31
 - program monitoring *vs.*, 203–204
 - program scope and, 100
 - providing information to help stakeholders do better, 23–24
 - as situational, 51
 - social justice and, 16
 - standards, 17
 - steps in, 17–18
 - systems view of, 3–6
 - theory-driven, 25–26
 - top-down approach, 388–389, 389–394
 - use of results, 214, 272
- Program evaluation approaches
 - choice of, 51–53
 - in comprehensive evaluation typology, 37–38(table), 41–44
- Program evaluation capacity, constructive assessment and building, 270–272
- Program evaluation design, 18–20
 - assertion of program's stage of development, 19
 - budget and timeline, 20
 - communicating about action plan for future, 119
 - evaluation types, approaches, and methodology, 20
 - logic model or program theory for describing program, 19
 - purposes and background information about intervention program, 18
- Program evaluation guidance
 - phasing in, 219–220
 - pilot testing, 219
 - principles for developing, 218–221
 - stakeholder input in developing, 215
- Program evaluation strategies, 37–38(table), 44–50
 - choice of, 51–53
 - in comprehensive evaluation typology, 37–38(table), 44–50
 - development strategies, 47–50
 - merit assessment strategies, 44, 46–47
- Program evaluation theories, reductionist-fluid complexity continuum and, 30–31
- Program evaluation typologies, 7–14
 - applications, 14
 - conclusive outcome evaluation, 12–13
 - conclusive process evaluation, 11–12
 - constructive outcome evaluation, 12
 - constructive process evaluation, 11
 - formative *vs.* summative evaluation, 7–9
 - fundamental, 10–13, 10(fig.)
 - hybrid evaluation types, 13–14
- Program implementation, accountability and, 180–181
- Program implementers
 - action model and, 76
 - action plan development and, 126–127
 - in anti-secondhand smoking program, 375
 - pilot testing and, 142, 145
 - program review/development meeting and, 169
 - theory-driven process evaluation of, 196–198
- Program improvement
 - evaluation and, 23–24
 - hybrid process evaluation and, 190
- Program model, 169
- Program monitoring
 - defined, 203–204
 - outcome monitoring, 206–208
 - process monitoring, 204–206
 - program evaluation *vs.*, 203–204
- Program-monitoring systems within organizations, 208–228
 - developing, 210–211
 - elements of, 209–210
- Program plan, defined, 97–98
- Program planning
 - action model/change model schema and, 92
 - comprehensive evaluation typology and, 37(table), 39–400, 42
 - incorporation of evaluation techniques into, 95
 - See also* Program scope
- Program review/development
 - meeting, 43, 49
 - advantages and disadvantages of, 167–168

- example, 169–171
- initial implementation evaluation
 - and, 165–171
- principles and procedures, 166–167
- purpose of, 165
- Program scope, 98
 - action plan and, 117–119, 118(fig.)
 - background information
 - provision strategy and approaches, 101–103
 - conceptual framework of, 98–100, 99(fig.)
 - conceptualization facilitation approach, 103–112
 - constructive process evaluation and modifying or clarifying, 177–179
 - garbage reduction program, 134, 135(fig.)
 - HIV-prevention program, 137–139, 138(fig.)
 - moving to action plan, 115
 - questions to inform evaluator's commentary on, 146–147
 - reasons to develop, 100
 - relevancy testing approach, 112–115
 - strategies for articulating, 100–101
- Program theory, 25–26, 42, 49, 65–66
 - building consensus among stakeholders regarding, 309
 - checking for potential vulnerabilities using, 274–275, 292–293
 - criticism of theory-driven outcome evaluation and, 338–339
 - defined, 66
 - evaluation design and, 19
 - identification of determinants and, 109
 - integrative process/outcome evaluation and, 329
 - stakeholders' implicit theories and, 368–369
 - theory-driven outcome evaluation and, 309–310
 - underlying anti-secondhand smoking program, 371–373
- Propriety, program evaluation and, 392
- Propriety standards, 17
- Providers Advocating for Sexual Health Initiative, theory-driven process evaluation of, 195–198
- Public area, as service setting, 126
- Public health programs, randomized controlled trials and efficacy of, 369
- Public sector, intervention programs in, 24
- Pure independent effects, 300
- Qualitative evaluation/qualitative methods
 - in HIV prevention program, 295
 - in holistic effectuality evaluation approach, 289, 295
 - in intervening mechanism evaluation, 318
 - in viability evaluations, 408
- Quality assurance, program implementers and, 127
- Quantitative design to address biases and assess change, 288–289
 - in HIV prevention program, 294–295
- Quantitative methods
 - in intervening mechanism evaluation, 318–319
 - in viability evaluations, 407–408
- Quantock, C., 160
- Quasi-experimental designs, 256–258
 - case illustration of, 257–258
 - interrupted time-series design, 257–258
 - nonequivalent comparison group design, 256–257
- Racism
 - differentiating plausible goals from window-dressing goals in program to eliminate, 240–241
 - establishing goals for media campaign against, 106, 107
- Randomized controlled trials (RCTs), 26, 252–253
 - assessing real-world effects and, 285
 - basis of validity and, 262
 - debates over as evaluative method, 260–262
 - effectual cogency and, 410
 - efficacy evaluations and, 392–393, 394
 - ethical and/or feasibility criteria and, 256
 - evaluation of needle exchange programs using, 403
 - evidence-based intervention movement and, 404
 - evidence produced by, 301
 - methodological criticisms of, 261
 - practical considerations, 262
 - real-world evaluation and, 263–264
 - top-down approach and, 389, 390
 - Zumba weight loss project and, 276–280
- Reality testing, 48
- Real world
 - evidence-based interventions and, 390–394
 - stakeholder theory-based interventions and, 370–371
- Real-world effectiveness, efficacious evidence *vs.*, 380–381
- Real-world joint effects, 300–301
- Real-world outcome evaluation, 14, 44, 260–302
 - building evidence from the ground up, 301–302
 - Campbellian validity typology and, 264–266
 - checklist for ranking real-world evaluations, 296–298, 297(table)
 - conclusive assessment, 269–270, 275–280
 - constructive assessment, 269–275
 - debates over experimentation evaluation approach, 260–262
 - efficacy *vs.* effectiveness evaluation and, 263–264
 - experimentation evaluation approach, 264–266, 299–301
 - holistic effectuality evaluation approach, 266–269, 298–301
 - incorporating adjuvants into, 268–269
 - methodology for, 284–286
 - randomized controlled trials and, 263–264
 - reasons for adjuvants for, 281–283
 - research steps for assessing real-world effects, 286–296

- Reasoning
 backward, 89, 90, 91
 forward, 90
 forward/backward, 91
- Reconnecting Youth program,
 effectiveness of, 390
- Recruitment
 pilot testing strategy for,
 144–145
 program implementer, 126–127
 target population, 131
- Reductionism, 26, 30–31
- Reeler, D., 343n
- Referral, reaching target population
 through, 131
- Referral fidelity evaluation,
 184–185
 example, 185
- Reichardt, C. S., 262, 289
- Reinforcement Alcohol Prevention
 Program, pre-experimental
 outcome evaluation of, 255
- Reinvention
 change in program plan, 164
 fidelity *vs.*, in conclusive process
 evaluation, 189–190
- Related organizations, 128
- Relevancy of formative
 evaluation, 155
- Relevancy testing, 42, 48, 112–115
 research example of, 113–115,
 114(fig.)
- Relevant goals, 232–233
- Renger, R., 344n
- Remedial action, formative and
 evaluation and choice of,
 157–158
- Research design, for theory-driven
 outcome evaluation, 310
- Research methods
 flexible application of, 86
 for formative evaluation,
 155–156
 for theory-driven process
 evaluation, 191
- Research questions, to inform action
 plan, 121
- Research steps, for assessing
 real-world effects, 286–296
 contextualizing intervention
 in real-world setting,
 286–288, 293
 example, 291–296
 replication of mesological
 intervention, 291
- using auxiliary design to
 triangulate evidence,
 289–291, 295–296
- using unobtrusive quantitative
 design to address biases and
 assess change, 288–289,
 294–295
- Residential program for mentally ill,
 macro-level contextual support
 for, 77
- Resolving Conflict Creatively
 Program, viable cogency
 and, 397–398
- Resources
 choosing implementing
 organization and, 124
 ecological context and, 77–78
- Responsiveness
 prioritizing in evaluation, 28
 synthesizing with scientific
 requirement, 28–29
- Retention, adjuvants and
 encouraging, 283
- Rigor of evaluation, 52–53
- Risk behavior, process monitoring
 and clients', 204
- Rossi, P. H., 318
- Rough draft, to facilitate
 discussion, 91
- Sabatier, P. A., 344n
- Sackett, D. L., 404, 405
- SAFER model, 344
- Samples, pilot testing, 142
- Sampling approach, 400
- Scale, comprehensive scanning
 and, 162
- Schneider, M., 320
- School attendance intervention
 program, creating SMART
 goals for, 233–234
- School-based antismoking campaign,
 intervening mechanism
 evaluation of, 316–317, 317(fig.)
- School-based health education
 program, efficacy evaluation
 of, 252–253
- School-based nutrition education
 program, intervention fidelity
 evaluation of, 184
- School/community center/club
 facility, as service setting, 126
- School dropout intervention
 programs, target population
 fidelity evaluation of, 188
- School violence prevention program,
 viable cogency and, 397–398
- Scientific credibility of program
 evaluations, 22–23
- Scientific methods, stakeholder
 requirement and, 27–28
- Scientific requirement
 prioritizing in evaluation, 28
 synthesizing with responsiveness
 requirement, 28–29
- Screening, 78
- Scriven, M., 7–9, 12, 16, 22, 23,
 238, 261, 336–337
- Secor, C., 159
- Selection threat to internal validity,
 250, 283
- Self-efficacy
 of anti-secondhand smoking
 program, 378
 determinants and, 72
 successful intervention and,
 382–383
- Self-selection
 in real-world evaluation, 286
 as threat to internal validity, 283
- Self-selection bias, 277–278
- Service delivery
 for anti-secondhand smoking
 program, 374–375
 constructive outcome assessment
 of, 271–272
 as moderator, 323–324, 324(fig.)
 modes of, 125
 program review/development
 meeting and, 170
- Service delivery fidelity evaluation,
 185–186
 example, 186
- Service delivery protocols
 action model and, 74–75
 action plan development and,
 124–126
 in pilot testing, 144
- Service intervention, implicit theory
 for, 307–308, 308(fig.)
- Service settings, 125–126
- Settings, service delivery protocol
 concern, 75
- Sex education in schools, identifying
 unintended outcomes, 243
- Sex worker outreach program,
 constructive assessment of, 274
- Shadish, W. R., 248, 262, 318
- Shapiro, J. P., 159
- Short-term goals, 106–107

- Single-entry evaluation, 55–56
- Site visits, in formative evaluation, 157
- 60 Minutes*, 11
- Skin cancer prevention, formative research and, 102–103
- Small-scale surveys, in formative evaluation, 155, 157
- SMART goals, 43, 231–234, 268, 270
 - attainable, 232
 - measurable, 232
 - putting SMART characteristics together in goals, 233–234
 - relevant, 232–233
 - specific, 232
 - time-bound, 233
- Smoker's hotline, target population fidelity evaluation of, 188–189
- Sobriety program, providing referrals or case management services, 133
- Social betterment programs
 - adjuvants and, 281–282
 - bottom-up approach and, 401–403
 - credible evidence of, 405
 - distinction between operational and official goals and, 241
 - integrated evaluation of, 27
 - top-down approach and, 389
- Social cognitive theory, 255, 379
- Social cognitive theory-based training, for community health advisors, 372
- Social connection adjuvant, 293–294
- Social justice, program evaluation and, 16
- Social learning theory, 72, 366
- Social science methodology, backward reasoning and, 90
- Sociodemographic variables
 - as moderators, 323, 323(fig.)
 - process monitoring and, 204
- Somera, D. M., 306–308
- Specific goals, 232
- Spencer, L. M., 395
- Spouse abuse intervention programs
 - descriptive assumptions and, 67
 - nonequivalent comparison group evaluation of, 257
 - prescriptive assumptions and, 68
 - relation between determinants and program outcomes, 108
 - relationships among components of action model, 79–80
- Stage-specific needs, 51
- Stake, Robert, 8, 9, 22
- Stakeholder credibility of program evaluations, 22–23
- Stakeholder requirement, 26–28
- Stakeholders, 16
 - action model/change model schema and, 86–91, 360–361
 - action plan and, 118
 - bottom-up approach and, 402, 409
 - building consensus among regarding program theory, 309
 - clarifying implicit theory, 306–309, 307(fig.), 308(fig.)
 - clarifying intentions, 181–182
 - clarifying stakeholders' theory, 87
 - communicating facts about evaluation strategy/approach, 53–54
 - delineation of strategy, 85–86
 - development of goals, constructive outcome evaluation and, 231
 - discussing findings of plausibility inquiry with, 244–245
 - engaging, 18
 - evaluability assessment and, 235
 - evaluations that help stakeholders do better, 23–24
 - feedback and, 5, 7
 - helping clarify a program plan (See Program scope)
 - integrated evaluation perspective and, 29
 - integrative cogency model and, 396–397
 - logic model engagement and, 60–61
 - matching evaluation strategy to needs of, 44–45
 - matching internal/external purposes of, 51–53
 - outcome evaluation and, 40–41
 - participating in constructive outcome assessment, 272–274
 - participating in developing program monitoring/evaluation system, 213–214, 223–226, 227
 - participatory modes for development facilitation, 88–89
 - preparing rough draft to facilitate discussion, 91
 - reasons for participation in evaluation, 16
 - reviewing existing documents and materials, 87
 - role in formative evaluation, 157–158
 - theorizing methods and background of, 104
 - theorizing procedures for development facilitation, 89–91
 - theory-driven process evaluation and, 190–191
 - understanding of theory-driven evaluation and, 309
 - use of process-monitoring data, 205
- Stakeholder theory, 72
 - choice of determinants and, 109
 - integrating with formal theory, 386
- Stakeholder theory-based interventions, 366–369
 - action theory success and, 383
 - evidence for merits of, 383–386
 - formal theory-based interventions *vs.*, 365–368
 - relative value of, 369–371
- Standards
 - accuracy, 17
 - feasibility, 17
 - program, 17
 - program evaluation, 392
 - proprietary, 17
 - utility, 17
- Stanford Chronic Disease Self-Management Program (CDSMP), 384
- Stanley, J., 248–249, 398
- Start-up programs, evaluation of, 39
- Static group comparison, 254
- Statistical conclusion validity, 248
- Statistical regression, 250
- Stickney, E. K., 161
- Stigmatization, as barrier to participation, 132
- Stokols, D., 321
- Store/shop interior, as service setting, 126
- Straus, M. A., 257
- Strecher, V. J., 113–115, 366
- Street outreach programs, target population fidelity evaluation of, 187

- Strength, fidelity evaluation and, 183
- Structural equation modeling, 319
- Students, use of book, 31–32
- Stufflebeam, D. L., 337–338
- Subjective nature of conclusions, evaluation and, 262
- Suchman, E., 247
- Suicide prevention program, identification of determinants and, 110
- Summative evaluation
 - defined, 8
 - formative evaluation *vs.*, 7–9
- Summit-style meetings, action plan development and, 130
- Support group, 125
- Surveys, 102, 408
- Systematic inquiry, 17
- Systematic scanning, 163–164
- Systematic Screening and Assessment Method, 386
- System change, 362–363
 - defined, 347
- System integration, integrative cogency model and, 396–397
- Systems view, of intervention programs and evaluation, 3–6, 4(fig.)
- Targeted generalization, 400
- Target population
 - action model and, 78–79
 - action plan and, 119
 - action plan development and, 130–133
 - anti-secondhand smoking program, 373
 - fidelity evaluation, 186–189
 - identifying, 105
 - pilot testing and, 144–145
 - program review/development meeting and analysis of, 169–170
 - relevancy testing approach and, 112–113
 - theory-driven process evaluation of, 196
- Technical assistance
 - development of program monitoring/evaluation systems and, 221–223, 227–228
 - program implementers and, 127
 - program-monitoring systems and, 209–210
- Technical expertise, choosing implementing organization and, 123
- Teen pregnancy prevention program, relevancy testing approach and, 113
- Telephoning, service delivery and, 125
- Testing threat to internal validity, 250, 282
- Theoretical justification for intervention, 111
- Theorizing methods, conceptualization facilitation approach and, 104–109
- Theorizing procedures for development facilitation, 89–91
- Theory-based evaluation, 312
- Theory-driven evaluation, 25–26
 - program theory and, 65
- Theory-driven outcome evaluation, 14, 44, 50, 304–338
 - advancing moderating mechanism models, 324–326, 325(fig.), 326(figs.)
 - clarifying stakeholders' implicit theory, 306–309
 - elements of, 310–311, 311(fig.)
 - guidelines for conducting, 309–310
 - integrative process/outcome evaluation approach, 327–333, 328(fig.), 330(fig.), 332(fig.)
 - intervening mechanism evaluation approach, 312–321, 312(fig.)
 - moderating mechanism evaluation approach, 321–327, 322(fig.), 323(fig.), 324(fig.)
 - overview, 304–305
 - reply to criticisms of, 336–338
 - types of, 310–312
 - unintended effects and, 333–336
- Theory-driven process evaluation, 43
 - anti-drug abuse program evaluation, 193–195, 194(table)
 - HIV-prevention intervention evaluation, 195–198
 - for mature implementation evaluation, 190–198
 - unintended effects and, 198
 - workplace smoking policy evaluation, 192–193
- Theory of change, 66
- Theory of change evaluation, 312
- Tilly, N., 305
- Time-bound goals, 233
- Timeline, program evaluation, 20
- Timeliness
 - of evaluation, 53
 - of formative evaluation, 155
- Titcomb, A., 344n
- Tobacco Use Prevention and Control Task Force, outcome monitoring of, 206–207
- Top-down approach
 - bottom-up approach *vs.*, 400–401, 401(fig.), 409–410
 - lessons learned from applying to program evaluation, 389–394
 - transitioning to dissemination and, 388–389
- Training
 - exit surveys, 377
 - of outreach workers, 377–378
 - of program implementers, 126–127
 - in program-monitoring/evaluation systems, 210, 221–222
- Transferability
 - credible evidence and, 406
 - efficacy and effectuality and, 405–406
 - in integrative cogency model, 395
- Transferability evaluation, 44, 399–400
- Transferable cogency, 398
- Transformation, program, 4
- Transient populations, outcome monitoring and, 207
- Transitional Aid Research Project, dynamic model of, 318
- Transportation, as barrier to participation, 132–133
- Treatment
 - in change model, 73–74
 - choosing, that affect determinant, 110–112
 - See also* Intervention
- Treatment integrity, 183
- Treatment programs, 3
- Trochim, W. M. K., 289, 290, 305
- Troubleshooting strategy, 47–49
 - for formative evaluation approach, 154–165
 - for implementation problems, 179–180

- for initial implementation evaluation, 153–171
 - for pilot-testing approach, 141–146
 - program review/development meeting and, 165–171
 - for relevancy testing approach, 112–115
- Truancy, changing determinant of, 111
- Turner, N. C., 350
- Tuskegee syphilis experiment, 22
- Tutorial program for children of Vietnamese and Laotian descent, identifying unintended outcomes, 243
- 20/20, 11
- Undercoverage, 131
- Unintended effects
 - field study detection of, 335–336
 - formal specification of, 334–335
 - theory-driven outcome evaluation and, 333–336
 - theory-driven process evaluation and, 198
- Unintended outcomes, methods for identifying, 243–244
- US Food and Drug Administration, 389
- United Way of America, logic model for, 59, 60(fig.)
- Urban, J. B., 406
- Utility, program evaluation and, 392
- Utility standards, 17
- UTOS approach, 400
- Validity
 - basis of, 262
 - construct, 248, 305
 - external, 248–249, 261, 262
 - statistical conclusion, 248
 - See also* Campbellian validity typology; Internal validity
- Validity-focused outcome evaluation, 44
 - experimentation evaluation approach, 247–258
 - foundation of, 247–248
- Variables
 - intervening, 71
 - mediating, 71
- Veterans Affairs scandal, disagreement on which goals to evaluate, 238–239
- Viability
 - credible evidence and, 406
 - efficacy and effectuality and, 405
 - integrative cogency model and, 395
 - outcome evaluation and, 381–382, 386
- Viability evaluation, 44, 399
 - cost savings and, 409
- Viable cogency, 396–398
 - focusing on, 407–408
- VOICES/VOCES, 384, 393
- Voluntary counseling and testing for HIV program, real-world outcome evaluation of, 291–296
- Volunteerism, implicit theory for, 307–308, 308(fig.)
- Volunteers, program implementation and, 127
- Waldo, G. P., 244
- Wandersman, A., 85
- Weber, Max, 69
- Web posting, service delivery and, 125
- Weinstein, R., 192
- Weiss, C. H., 15, 65, 66
- Welfare program
 - identifying unintended outcomes of, 244
 - integrative process/outcome evaluation of, 331–333, 332(fig.)
- Welfare reform
 - goal of, 71
 - micro-level contextual support and, 77
- Wholey, J. S., 58, 59, 235, 236
- Winkleby, M. A., 121
- Working group, 88–89
 - clarifying stakeholders' implicit theory and, 308–309
 - conceptualization facilitation and, 103–104, 122
 - mature implementation evaluation and, 178
- Workplace nutrition program, intervening mechanism evaluation of, 315, 316(fig.)
- Workplace smoking policy
 - field study detection of unintended effects of, 335–336
 - theory-driven process evaluation of, 192–193
- Work-release program, identifying unintended outcomes of, 244
- YMCA drop-in center, concept mapping of, 178–179
- Youth crime reduction program. *See* Juvenile crime reduction/prevention programs
- Youth Risk Behavior Survey, 207
- Zald, M. N., 242
- Zumba weight loss project, conclusive assessment of, 275–280, 281, 282, 283, 284–285, 286–287, 300–301