

PETER H. ROSSI
HOWARD E. FREEMAN

EVALUATION

A SYSTEMATIC
APPROACH



PETER H. ROSSI
HOWARD E. FREEMAN

EVALUATION

A SYSTEMATIC
APPROACH

5



SAGE Publications

International Educational and Professional Publisher

Newbury Park London New Delhi

Copyright © 1993, 1989, 1985, 1982, 1979 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.

For information address:



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

SAGE Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Rossi, Peter Henry, 1921-

Evaluation: a systematic approach / Peter H. Rossi, Howard E. Freeman.—5th ed.

p. cm.

Includes bibliographical references (p.) and index.

ISBN 0-8039-4458-6

1. Evaluation research (Social action programs) I. Freeman, Howard E. II. Title.

H62.R666 1993

361.6'1'072—dc20

92-37082

98 99 00 01 12 11 10 9 8

Sage Production Editor: Astrid Virding

To the Memory of
Howard E. Freeman
Pioneer in Evaluation Research and
Prolific Contributor in Practice and Theory
and
Dearest Comrade

Preface

Evaluation: A Systematic Approach provides an introduction to the broad set of research activities essential for designing, implementing, and appraising the utility of social programs. At both the local and national levels, and in both industrialized and less developed countries, evaluators are engaged in developing and testing innovative initiatives designed to ameliorate and control social problems and in refining and assessing the effectiveness and efficiency of programs already in place. The work of evaluation researchers influences social policies and guides efforts to improve the lives and social milieus of community members.

Since we embarked on the first edition of *Evaluation* well over a decade ago, it has been our ambition to communicate the technical knowledge and professional experience we have gained doing evaluations. We remain committed to this aspiration, for we continue to believe that evaluation research is a dynamic, exciting, and socially valuable genre of applied social research. Accordingly, we have revised, updated, and augmented this textbook as the field of evaluation and we as practitioners have changed.

Students, practitioners, novice social researchers, public officials, sponsors of social programs, and the legendary intelligent laypersons are the intended audiences for this book. Although experienced and skillful researchers may find it elementary in spots, we hope it will be useful to them and particularly to their fledgling associates as a reference volume. For all readers, we believe it can increase technical competencies, broaden outlooks on the field, and identify sources that can amplify the ideas we discuss. We attempt to convey the material within a coherent framework, with ample attention to the political constraints and interpersonal dynamics that must be taken into account in order to do successful evaluations and to appreciate the work being undertaken by others.

This book has a cosmopolitan past that we hope still influences its form and content. Its beginnings go back to the mid-1970s, when we presented independently written, but remarkably similar, papers on the state of evaluation research at a UNESCO conference (Freeman, 1977; Rossi and Wright, 1977). Staff of the Organization for Economic Cooperation and Development (OECD) who attended the conference urged us to expand our papers into a survey of evaluation methods that could be employed in less developed nations. After being revised at two working conferences in Paris, that book was finally published in English and French (Freeman, Rossi, and Wright, 1980).

Our experience in writing the OECD volume inspired us to write a "textbook" on evaluation research as it is practiced in the United States. The OECD book and *Evaluation: A Systematic Approach* are siblings but not clones. Unlike the OECD volume, this book presumes that readers have at least undergraduate and generally some graduate social research training, that in their work there is reasonable access to computers and to libraries, and that peers with similar interests are readily available.

When the first edition of *Evaluation* was published, we were rather smug in feeling that we had provided evaluators and those who use their findings with a reasonably definitive text covering all aspects of the field. After all, between us we had more than eight decades of experience doing applied research and had undertaken, consulted on, and criticized evaluations in a large number of social program areas. The book was well received, used in both undergraduate and graduate courses in a number of social science disciplines and professional schools, and bought—and, we presume, read—by a fairly large audience.

Since then both the book's success and its shortcomings, together with the rapid changes that have taken place in the field, have prompted three revisions, the last published in 1989. In addition to further revising the conceptual material and refining the technical discussions contained in earlier editions, in the third edition we had taken into account the increased evaluation activities among economists, the changing political climate, and, of course, the widespread adoption of microcomputers. In the fourth edition, we clarified our typology of evaluations and tried to set straight some of the confusion between designs. Also, we generally updated the rest of the book. Perhaps the most useful revisions were in the chapter on measuring efficacy. With Ann Witte's advice, we were able to improve it dramatically in the fourth edition—at least we think so.

Someone had suggested to us that authors should never write prefaces until their book is ready to "go to bed." By the time we

finished the revisions to the fourth edition, given the spirited and often raucous interaction that has characterized our many years of collaboration, we pledged never to revise *Evaluation* again. However, our own self-review and the responses of helpful colleagues have made another attempt to improve and update the text irresistible.

Although this fifth edition is not radically different from its predecessor, there are some notable changes. This edition provides a considerably revised chapter on program monitoring. Its new title, “Program Monitoring for Evaluation and Management,” is a strong clue to the new content of the chapter. In addition, we go much more deeply into meta-analysis, particularly with regard to its use in the design and modification of interventions.

Perhaps most noticeably, we have added a tenth chapter to this edition. In previous editions, we had a chapter that introduced the range of procedures for evaluating program impact, moved on to describing “true experiments” in another chapter, and then considered all other procedures in a third. In this edition, we have separated the material on non-random designs (most of these designs are referred to in the field as “quasi-experiments”) from procedures used to assess full-coverage programs (that is, ones where comparison groups of any sort are usually impossible to construct). We did so both because of the differences in outlook required for the evaluation of full-coverage programs and in order to be able to devote more space to this group of procedures.

In addition to these refinements, we have added new exhibits and deleted some of the ones included in earlier editions, and updated the reference section. In selecting new exhibits we have tried to correct somewhat our bias of choosing “national” evaluations and neglecting local and state ones.

While we have learned not to seek perfection, we believe that these and a number of smaller changes nudge the book a bit closer to this lofty goal. Finally, John Bergez’s editing, again, has made this edition into a much better book than it would have been if our original manuscript had been set in type.

Instructors and others should note that a *Workbook for Evaluation: A Systematic Approach* is available to accompany this text. The workbook includes brief reviews of each chapter, review questions, design and analysis tasks presenting case studies for readers to tackle, and additional readings. The new edition of the workbook also includes a complete bibliography of journals featuring evaluation studies and issues.

A personal note to conclude this preface: The work, conflict, and fun that went into this volume were shared equally by the two

authors. Age and notoriety were the major determinants of the order of authorship. Rotating authorship in alternate editions would be confusing; besides, why fix something that works? Despite the impossibility of writing a definitive text, we are very proud of this book, and particularly of the fact that it has been used in hundreds of classes, both graduate and undergraduate, and in a wide range of social science departments and professional schools. Perhaps most important, over the years collaborating on it has turned an acquaintanceship into an enduring friendship. We are perhaps most grateful for this by-product of our venture.

P. H. R. and H. E. F.

Postscript

Within a few hours after we finished the index to this edition, Howard Freeman died while on the plane going back to his home in Los Angeles. In that terrible event, evaluation research lost one of its pioneers, its best practitioner, and a prolific contributor. I lost a true friend and the grandest of collaborators. Evaluation will miss his lively presence. I will miss his quick intelligence, sharp wit, and warm friendship.

I dedicate this book to his memory.

Peter H. Rossi
October 1992

Contents

Preface	ix
1 Programs, Policies, and Evaluations	3
♦ What Is Evaluation Research?	5
♦ A Brief History of Evaluation Research	9
♦ Evaluation Research in Practice	27
♦ An Overview of Evaluations	34
♦ How Evaluations Are Used	46
♦ Who Can Do Evaluations?	52
♦ Summary	54
2 Diagnostic Procedures	57
♦ The Role of Evaluators in Diagnosing Social Conditions and Needs	58
♦ Identifying and Defining Social Problems	60
♦ Specifying the Problem: Where Is the Problem and How Big Is It?	63
♦ Defining and Identifying the Targets of Interventions	85
♦ Useful Concepts in Target Definition	96
♦ Summary	103
3 Tailoring Evaluations	105
♦ A Note on Stakeholders	110
♦ Evaluation Concerns in the Planning of New Programs	111
♦ Evaluating Established Programs	142
♦ Fine-Tuning Established Programs	154
♦ Linking Evaluations to Programs	158
♦ Summary	159

4	Program Monitoring for Evaluation and Management	163
♦	What Is Monitoring?	164
♦	Perspectives on Program Monitoring	166
♦	Monitoring Target Participation: Program Coverage and Bias	175
♦	Monitoring Delivery of Services	187
♦	Collecting Data for Monitoring	197
♦	Analysis of Monitoring Data	209
♦	Feedback from Monitoring	211
♦	Summary	212
5	Strategies for Impact Assessment	215
♦	Key Concepts in Impact Assessment	217
♦	Extraneous Confounding Factors	222
♦	Design Effects	226
♦	Design Strategies for Isolating the Effects of Extraneous Factors	240
♦	A Catalog of Impact Assessment Designs	243
♦	Judgmental Approaches to Impact Assessment	252
♦	Pooling Evaluations: Meta-Analysis	253
♦	Quantitative versus Qualitative Data in Impact Assessments	254
♦	Inference Validity Issues in Impact Assessment	255
♦	Choosing the Right Impact Assessment Strategy	258
♦	Summary	258
6	Randomized Designs for Impact Assessment	261
♦	Units of Analysis	261
♦	Experiments as an Impact Assessment Strategy	262
♦	Analyzing Randomized Experiments	278
♦	Limitations on the Use of Randomized Experiments	290
♦	Summary	294
7	Quasi-Experimental Impact Assessments	297
♦	Quasi-Experimental Impact Assessment	298
♦	Constructing Control Groups by Matching	303
♦	Equating Groups by Statistical Procedures	311
♦	Regression-Discontinuity Designs	319
♦	Generic Controls	320
♦	Some Cautions in Using Constructed Controls	322
♦	Summary	329

8	Assessment of Full-Coverage Programs	333
♦	Non-Uniform Full-Coverage Programs	335
♦	Reflexive Controls	340
♦	Shadow Controls	351
♦	Summary	361
9	Measuring Efficiency	363
♦	Key Concepts in Efficiency Analysis	365
♦	Methodology of Cost-Benefit Analysis	376
♦	Cost-Effectiveness Analysis	397
♦	Summary	400
10	The Social Context of Evaluation	403
♦	The Purposefulness of Evaluation Activities	405
♦	The Social Ecology of Evaluations	406
♦	The Profession of Evaluation	430
♦	Utilization of Evaluation Results	443
♦	Epilogue	453
♦	Summary	454
	References	456
	Author Index	475
	Subject Index	480
	About the Authors	487

EVALUATION

KEY CONCEPTS FOR CHAPTER 1

Comprehensive evaluation:	Research and analysis covering the conceptualization and design of interventions, the monitoring of program interventions, and the assessment of program utility.
Conceptualization and design analyses:	Studies of (1) the extent and location of target problems for intervention, (2) ways to operationally define target populations, and (3) the suitability of the proposed intervention.
Cost-benefit analyses:	Studies of the relationship between project costs and outcomes, with both costs and outcomes expressed in monetary terms.
Cost-effectiveness analyses:	Studies of the relationship between project costs and outcomes, expressed as costs per unit of outcome achieved.
Delivery system:	Organizational arrangements needed to provide program services, including staff, procedures and activities, physical plants, and materials.
Efficiency evaluations:	Analyses of the costs (inputs) of programs in comparison to either their benefits or to their effectiveness (outputs). Benefits, like inputs, are expressed in monetary values; effectiveness is measured in terms of substantive changes in the behavior or conditions that are the foci of a social program.
Fine-tuning evaluations:	Evaluations undertaken to examine the implementation, effectiveness, and efficiency of modifications in existing programs.
Impact assessment:	Evaluation of whether and to what extent a program causes changes in the desired direction among a target population.
Intervention:	Any program or other planned effort designed to produce changes in a target population.
Monitoring:	Assessing the extent to which a program is (1) undertaken consistent with its design or implementation plan and (2) directed at the appropriate target population.
Stakeholders:	Individuals or organizations directly or indirectly affected by the implementation and results of social programs.
Target population:	The persons, households, organizations, communities, or other units at which an intervention is directed.
Target problem:	The conditions, deficiencies, or defects at which an intervention is directed.
Utility assessment:	Study of the effectiveness (impact) or efficiency (costs-to-benefits ratios or cost effectiveness) of programs.

Programs, Policies, and Evaluations

This chapter introduces evaluation research as a robust arena of activity directed at collecting, analyzing, and interpreting information on the need for, implementation of, and effectiveness and efficiency of intervention efforts to better the lot of humankind. Systematic evaluations of both existing and new social programs are now commonplace. Evaluations are undertaken for a variety of reasons: to judge the worth of ongoing programs and to estimate the usefulness of attempts to improve them; to assess the utility of new programs and initiatives; to increase the effectiveness of program management and administration; and to satisfy the accountability requirements of program sponsors. Evaluations also may contribute to substantive and methodological social science knowledge.

Evaluations serve different purposes, and call for different strategies, at various stages in the life of programs. In the planning stages of social intervention programs, evaluations focus on assessing the extent and severity of the problems requiring social intervention and on designing programs to ameliorate them. In the conduct of ongoing and new programs, evaluations help to determine the degree to which the programs are effective—that is, how successfully they are providing their intended target populations with the resources, services, and benefits envisioned by their sponsors and designers. In addition, impact assessments estimate the effects of the intervention. For accountability purposes, and to aid in decisions concerning whether programs should be continued, expanded, or curtailed, evaluations consider costs in relation to benefits and compare an intervention's cost effectiveness with that of alternative strategies for allocating scarce resources.

Some evaluations are comprehensive and consider all of these questions; others are directed at only some of them. In all cases, the aim is to provide the most valid and reliable findings possible within political and ethical constraints and the limitations imposed by time, money, and human resources.

Long before Sir Thomas More coined the word *utopia* in 1516, persons everywhere had been seeking a perfect world. That their aspirations, and ours, have hardly been realized is evident in the social problems and attendant personal problems that confront every country in the world. True, how we define social problems and estimate their scope, and which problems are salient to us, have changed over time with shifts in values and lifestyles. And it is equally true that communities, societies, and cultures differ widely in the attention they pay to particular problems. But now, as ever, to borrow from Charles Dickens, for some of us these are the best of times and for others the worst of times.

Since antiquity, persons and groups have sought to describe, understand, and ameliorate the defects in the human condition in myriad ways. As pretentious as it may sound, this book is rooted in an approach that has aspired to be effective in improving the quality of our physical and social environments and enhancing the chances of our individual and collective survival: the application of scientific procedures to social problems. If the term *evaluation research* is a relatively recent invention, the activities that we will consider under this rubric are not. They can be traced to the very beginnings of modern science in the 1600s. As Cronbach and his colleagues (1980) point out, Thomas Hobbes and his contemporaries four centuries ago were concerned with devising numerical measures to assess social conditions and to identify the causes of mortality, morbidity, and social disorganization. And certainly social experiments, the highest expression of contemporary evaluation research, are hardly a recent invention.

Indeed, one of the earliest “social experiments” took place in the 1700s. A British ship’s captain observed the lack of scurvy among sailors serving on the naval ships of Mediterranean countries. He noticed, too, that citrus fruit was a part of their rations. Thereupon he made half his crew consume limes while the other half continued with their regular diet. Despite much grumbling among the crew in the “treatment” half, the experiment was a success; it showed that consuming limes could prevent scurvy.

The good captain probably did not have an explicit “impact model” (a term we will discuss later in detail)—namely, that scurvy is a consequence of a vitamin C deficiency and that limes are rich in vitamin C. Nevertheless, the treatment worked, and British seamen eventually were compelled to consume citrus fruit regularly, a practice that gave rise to the still-popular label “limeys.” Incidentally, it took about fifty years before the captain’s “social program” was widely adopted. Then, as now, diffusion and acceptance of evaluation findings did not come easily.

❖ What Is Evaluation Research?

We begin this volume with a simple definition of evaluation, or evaluation research (we will use the terms interchangeably): *Evaluation research is the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs.* In other words, evaluation researchers (evaluators) use social research methodologies to judge and improve the ways in which human services policies and programs are conducted, from the earliest stages of defining and designing programs through their development and implementation.

At various times policymakers, funding organizations, planners, and program managers' staffs need to distinguish useful current programs from ineffective and inefficient ones, and to plan, design, and implement new efforts that effectively and efficiently have the desired impact on community members and their environments. To do so they must obtain answers to a range of questions. For example:

- ◆ What is the nature and scope of the problem requiring new, expanded, or modified social programs? Where is it located, and whom does it affect?
- ◆ What feasible interventions are likely to ameliorate the problem significantly?
- ◆ What are the appropriate target populations for a particular intervention?
- ◆ Is the intervention reaching its target population?
- ◆ Is the intervention being implemented in the ways envisioned?
- ◆ Is it effective?
- ◆ How much does it cost?
- ◆ What are its costs relative to its effectiveness and benefits?

Answers to such questions are necessary not only for broad, complex programs, such as nationwide health care or income maintenance efforts, but also for local, specialized projects, such as those offering occupational training in a small town, testing whether a new method of teaching mathematics will work, or increasing public safety in a large city. Providing those answers is the work of persons in the evaluation research field.

Although this text emphasizes the application of social research procedures to the evaluation of health, education, welfare, and other human service policies and programs, evaluation research is not restricted to such programs. Indeed, the Program Evaluation Methodology Division of the General Accounting Office has evaluated programs in a variety of fields, including the procurement and testing of military hardware, research programs in the repair of major highways, and other areas far afield from social interventions.

The approaches considered in this text also have utility in many other spheres of activity in which issues of efficiency, effectiveness, and impact are raised. For example, the mass communication and advertising industries use fundamentally the same approaches in developing media programs and marketing products; commercial and industrial corporations evaluate the procedures they use in selecting and promoting employees and organizing their work forces; political candidates develop their campaigns by evaluating the voter appeal of different strategies; consumer products are subjected to market testing; and administrators in both the public and private sectors are continually assessing the clerical, fiscal, and interpersonal practices of their organizations.

The distinction between these uses of evaluation lies primarily in the intent of the effort to be evaluated. Our emphasis in this text is on evaluation of programs designed to benefit the human condition, rather than those designed for other purposes, such as increasing profits or amassing influence and power. This choice of focus stems not from a sense of righteousness about the proper application of social research methods but rather from a heuristic need to limit the scope of the book.

Regardless of the type of social intervention under study, evaluations are systematic to the extent that they employ social research approaches to gathering valid, reliable evidence. This commitment to the “rules” of social research is at the core of our perspective on evaluation. This is not to say, however, that evaluation studies follow some particular social research style or combination of styles. Indeed, one of the distinguishing characteristics of program evaluation is that its methods cover the gamut of prevailing social research paradigms.

The importance of evaluating social programs—both those programs currently in effect and those in various stages of design and pilot testing—should not be underestimated. The continuing challenge of devising ways to remedy the serious deficiencies in the quality of human life, both in industrialized countries and in less developed nations, needs no elaboration here. But along with the need for purposeful, deliberate, and organized efforts to implement creative new initiatives and improve existing ones comes the need for evaluation. True, resource restraints in every country, including the United States, constrain social intervention activities. Moreover, the experiences of the past several decades have brought about a more realistic perspective on the barriers to successful implementation of social programs and the magnitude of outcome that can be expected from them (Weick, 1984). In general, social programs yield small gains, often to the chagrin of the those who advocate them (see Exhibit 1-A).