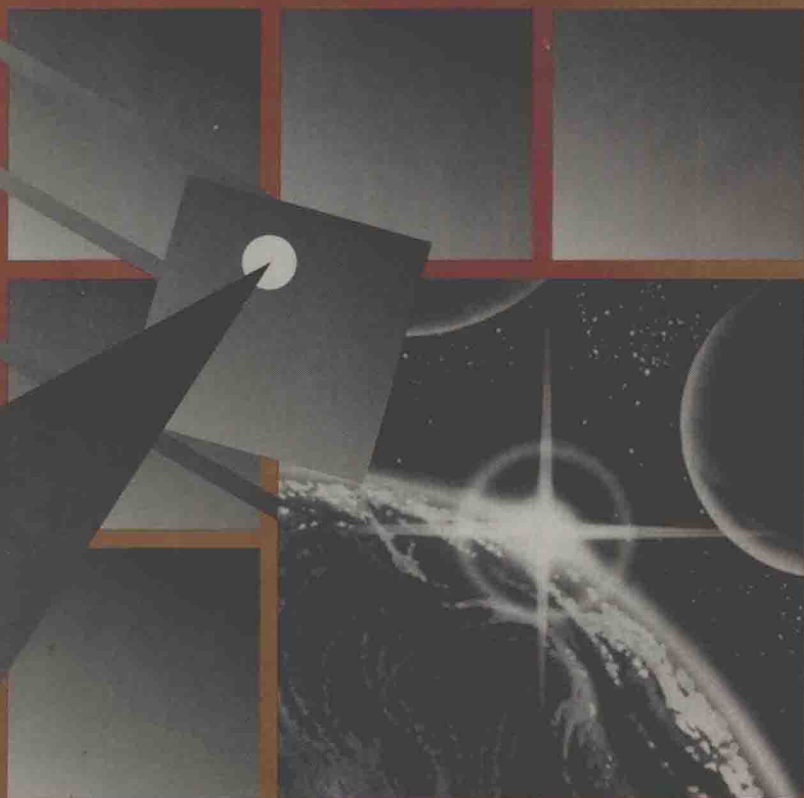


EVALUATION FOR THE 21st CENTURY

A HANDBOOK



EDITORS
ELEANOR CHELIMSKY
WILLIAM R. SHADISH

EVALUATION FOR THE 21st CENTURY

A HANDBOOK

EDITORS

ELEANOR CHELIMSKY

WILLIAM R. SHADISH



SAGE Publications

International Educational and Professional Publisher

Thousand Oaks London New Delhi

Copyright © 1997 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
Thousand Oaks, 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

Main entry under title:

Evaluation for the 21st century: A handbook / editors, Eleanor
Chelimsky and William R. Shadish.

p. cm.

Includes bibliographical references and index.

ISBN 0-7619-0610-X (acid-free paper). – ISBN 0-7619-0611-8 (pbk.:
acid-free paper)

I. Evaluation research (Social action programs). I. Chelimsky,
Eleanor. II. Shadish, William R.

H62.E8473 1997

001.4–dc20

96-25373

97 98 99 00 01 02 03 10 9 8 7 6 5 4 3 2

<i>Acquiring Editor:</i>	C. Deborah Laughton
<i>Editorial Assistant:</i>	Eileen Carr
<i>Production Editor:</i>	Astrid Viriding
<i>Production Assistant:</i>	Denise Santoyo
<i>Typesetter/Designer:</i>	Christina Hill
<i>Indexer:</i>	Teri Greenberg
<i>Cover Designer:</i>	Lesa Valdez
<i>Print Buyer:</i>	Anna Chin

In Memoriam

As this book goes to press, we are deeply saddened to hear of the passing of Donald T. Campbell in May 1996. We publish here the last paragraph of some remarks made about his work at an evaluation conference 5 years ago that remain entirely pertinent:

“Looking back over Campbell’s work, it is clear that he has consistently tried to integrate opposite poles in evaluative thinking, to bridge the gaps between them. This effort of his was crucial to evaluation’s survival over the past decade, when it was under attack, and it promises Campbell an enduring place, not only in the annals of applied social science research but also in whatever pantheon we have in America for people who can understand both sides of an issue and integrate them into a larger—but always elegant—framework.”

Preface

For the past 30 years, we have witnessed the development of a discipline and a profession of evaluation. Less than 4 years from now, we will be witnessing the passing of the millennium, which leads naturally to the question, What will evaluation be like in the 21st century? Our crystal ball is undoubtedly fallible, but in this book we present some of the features of the professional landscape that we suspect will become increasingly salient as the end of the 20th century comes and goes. We call this a handbook because it is our hope that as evaluators begin to encounter these developing features, they will recognize them from our description, be able to place them into context, and perhaps even return to read about them in more detail once these features impinge more strongly on their professional practice.

Evaluation is becoming increasingly international, but in ways that go beyond previous conceptions of what *international* means. *International* is no longer used only to describe the efforts of particular evaluators in individual countries around the world—although it certainly is still used this way, as in the descriptions of evaluation in China and in Denmark included in this book. Today, evaluation is also becoming international in the sense of being at the same time more indigenous, more global, and more transnational. By *indigenous*, we mean that evaluators in different countries around the world are developing their own infra-

structures to support their endeavors as well as their own preferred theoretical and methodological approaches. By *global*, we mean that developments in one part of the globe frequently affect people, institutions, and programs all around the world. An example in this volume is the description of how the evaluation of community development in a small town in Japan is influenced by policies and people in and from other nations. By *transnational*, we mean that the problems and programs that we are called upon to evaluate today often extend beyond the boundaries of any one nation, any one continent, or even one hemisphere. As is noted in Chapter 1, these include problems of pollution, of the economics of developing countries, and of the role of women in society. We cannot say exactly what the best responses to these internationalizing developments will be for evaluators, but we do know that recognizing the developments is the first step toward accommodating to them.

Evaluation is becoming increasingly diverse in what it is asked to evaluate. Michael Scriven once said that it is possible to evaluate anything (including evaluation itself). That observation looks prophetic in the context of the contents of this book. Today, the diverse array of things being evaluated is nearly bewildering—not just programs, personnel, and products, not only medical and military technology, but also foreign aid, defense policy, environmental interventions, nuclear power plant accidents, massive human rights violations, and World Bank loan programs. This diversity will undoubtedly challenge evaluators to develop new methods, new organizational strategies, and new conceptualizations of what they can and should be doing.

Evaluation is becoming increasingly self-conscious about its own identity. To be sure, we will continue to ask the question, What is evaluation? The now-classic answer is that evaluation is about determining merit or worth—although it is far from clear that we have consensus on that answer across the diversity of the evaluation profession. The more interesting development may be the increasing frequency with which we are asking the question, What makes evaluation different from other disciplines? For example, the increasingly close ties between the evaluation profession and the auditing profession prompt many thoughtful questions about the differences and similarities between the two that are explored in this book. Certainly we have much to learn from each other. Similarly, evaluators are asking more subtle questions about exactly which activities have priority in evaluation, under what circumstances,

for what purposes. Examples in this book concern the relative priority assigned to advocacy versus truth in evaluation and to evaluating programs versus empowering people to evaluate their own programs.

Evaluation continues to become ever more methodologically diverse. It is by now well established that the full array of social science methods belongs in the evaluator's methodological tool kit—tools from psychology, statistics, education, sociology, political science, anthropology, and economics. Clearly we must also include the logical tools given to us by philosophers who study the nature of evaluation. Further, evaluators continue to invent interesting constellations of methods, or strategies for approaching evaluation, such as cluster evaluation and empowerment evaluation. But even evaluators who accept all this diversity may be surprised by the tools they may still need to learn about, to judge from the chapters in this book—the methods of auditors who have extended their work from auditing financial records to auditing program performance, of evaluators who synthesize studies done quite differently, and perhaps even of physical scientists whose global questions make locally constrained methods impractical.

It is often uncomfortable to stir oneself from familiar cultural, topical, conceptual, and methodological niches. But if there is a moral to the present book, it is this: Evaluation in the next century will be far more diverse than it is today. So we must face the discomfort of stirring ourselves if we are to avoid being left behind.

Finally, evaluators, in whatever field of evaluation they may be, are likely to find themselves, at least sometimes, at odds with the political actors, systems, and processes in their countries that militate against the free flow of information required by evaluation. This means that as the world becomes more politically diverse and complex in the 21st century, evaluators will be called upon to exhibit considerable courage in the normal pursuit of their work. Because they are examples of admirable courage in using research to speak truth to power, we want to dedicate this book to six men of the 20th century: Vaclav Havel, C. Everett Koop, Jean Monnet, Daniel Patrick Moynihan, Gunnar Myrdal, and Elliot Richardson. Their strength and vision have paved the way for the next generations.

Eleanor Chelimsky
William R. Shadish

Contents

PREFACE	<i>xi</i>
1. The Coming Transformations in Evaluation <i>Eleanor Chelimsky</i>	1
<hr/>	
PART I: EVALUATION—YESTERDAY AND TODAY	27
2. Lessons Learned in Evaluation Over the Past 25 Years <i>Thomas D. Cook</i>	30
3. The Political Environment of Evaluation and What It Means for the Development of the Field <i>Eleanor Chelimsky</i>	53
<hr/>	
PART II: AUDITING AND EVALUATION	69
4. Evaluation as an Essential Component of “Value-for-Money” <i>L. Denis Desautels</i>	72
5. Auditing and Evaluation in Sweden <i>Inga-Britt Ahlenius</i>	80

6. Performance Auditing: Travelers' Tales <i>Christopher Pollitt and Hilikka Summa</i>	86
7. Evaluation and Auditing in State Legislatures: Meeting the Client's Needs <i>Roger A. Brooks</i>	109

PART III: PERFORMANCE MEASUREMENT AND EVALUATION	121
---	------------

8. Trends in Performance Measurement: Challenges for Evaluators <i>Joseph S. Wholey</i>	124
9. Performance Measurement in the United Kingdom (1985–1995) <i>Caroline Mawhood</i>	134

PART IV: INTERNATIONAL EVALUATION	145
--	------------

10. Evaluating Global Issues in a Community Setting <i>Masafumi Nagao</i>	149
11. The Development of Evaluation in the People's Republic of China <i>Hong Hongqi and Ray C. Rist</i>	170
12. Critical Comments on Evaluation Research in Denmark <i>Finn Hansson</i>	177
13. Evaluation, Markets, and Institutions in the Reform Agenda of Developing Countries <i>Eduardo Wiesner</i>	189
14. Evaluation in the World Bank: Antecedents, Instruments, and Concepts <i>Robert Picciotto</i>	201

PART V: NEW TOPICS FOR EVALUATION **215**

15. Evaluating Human Rights Violations 221
Ignacio Cano
16. Lessons of Immigration Policy and the Role of Research 234
John Nieuwenhuysen
17. Tracing Gender Issues Through Institutional Change
and Program Implementation at the World Bank 251
Gosette Murphy
18. Understanding the Impact of Development Projects on Women:
The Tunisia Institutional Development Fund Project 260
Michael Bamberger
19. The Analysis and Evaluation of Foreign Aid 272
Kristina Stenstrom
20. Evaluating the U.S. Nuclear Triad 284
Kwai-Cheung Chan and Jonathan R. Tumin
21. Evaluation, Nuclear Power Plant Remediation and
Redesign, and Russian Policy Making After Chernobyl 299
Vladimir Novikov
22. The Independent Evaluation of the Global Environment
Facility Pilot Phase 311
W. Haven North
23. Environmental Changes and Their Measurement:
What Data Should We Collect and What Collaborative
Systems Do We Need for Linking Knowledge to Action? 329
Charles A. Zrakat and William Clark

**PART VI: A SAMPLER OF THE CURRENT
METHODOLOGICAL TOOL KIT** **337**

24. Multimethod Evaluations: Using Case Studies Together
With Other Methods 344
Lois-ellin Datta

25. Cross-Design Synthesis: Concept and Application <i>Judith A. Droitcour</i>	360
26. Research Synthesis for Public Health Policy: Experience of the Institute of Medicine <i>Michael A. Stoto</i>	373
27. Empowerment Evaluation and Accreditation in Higher Education <i>David M. Fetterman</i>	381
28. Cluster Evaluation <i>James R. Sanders</i>	396
29. An Introduction to Scientific Realist Evaluation <i>Ray Dawson and Nick Tilley</i>	405
30. Single-Case Evaluation in British Social Services <i>Mansoor A. F. Kazi</i>	419
31. Twenty-One Years Old and Counting: The Interrupted Time Series Comes of Age <i>Robert G. Orrin</i>	443
<hr/>	
PART VII: AN ENDURING ARGUMENT ABOUT THE PURPOSE OF EVALUATION	467
32. Advocacy in Evaluation: A Necessary Evil? <i>Robert E. Stake</i>	470
33. Truth and Objectivity in Evaluation <i>Michael Scriven</i>	477
INDEX	501
ABOUT THE EDITORS	521
ABOUT THE CONTRIBUTORS	523

1

The Coming Transformations in Evaluation

Eleanor Chelimsky

The joint International Evaluation Conference held in Vancouver, Canada, during the first week of November 1995 was, in one participant's words, "a watershed event."¹ It marked the first time that five evaluation associations and more than 1,600 evaluators from 66 countries and five continents had come together.² The purpose was to take stock of where evaluation is today and where it is likely to be going in the next century.

The conference focused on three topics intended to reflect international experience in evaluation to date: first, global issues (centering on transnational problems rather than national or local ones); second, lessons learned from evaluation theory and practice of the past; and third, evaluation's changing characteristics (especially changes related to the adaptation of traditional methods and the use of nontraditional methods in different parts of the world).

There were 363 panels, workshops, and other sessions (each with a number of presenters); there were five plenary speakers; there were

preconference workshops focusing on a variety of evaluation methods; and throughout the conference—in the corridors and over coffee—spontaneous debates arose, inspired by particular sets of papers, perhaps, but bringing to bear a wonderful variety of evaluative experiences from far-flung places.

This book, of necessity, can present only selected papers from the conference and therefore gives a somewhat attenuated reflection of its extraordinary richness, spirit, and vitality. The hope is that what it must lose in diversity it will gain in focus as the scope is narrowed to just one subject—current changes in evaluation. In this chapter, I will discuss four themes that are related to those changes:

- the new political and policy context as we move into the 21st century
- new actors, topics, and nations in the field of evaluation
- new versions of old arguments about the purposes of evaluation and the implications of those purposes for methodology, for the use of findings, and for the role of the evaluator
- new and old lessons from the continuing evaluation experience

Although these are certainly not the unique themes of the chapters in this volume, they are very important ones. In this chapter, then, I will try to confront, as objectively as possible, the principal elements and controversies with respect to these themes, using the 33 chapters in this book as data for the discussion.

■ *A New Political and Policy Context for Evaluation*

The mere fact of examining the evaluative questions confronting, say, continents rather than nations, and facing outward toward worldwide challenges rather than inward toward domestic problems, brings three issues immediately to the fore that have not been traditional ones in evaluation.

First, evaluation must now deal with a number of global political forces that are transforming all societies—developing and developed nations alike—and at fairly high rates of speed. One of these forces is certainly the worldwide adoption of new technologies—for example,

automated workplaces or genetically engineered crops—with all the changes they imply for the old factory- and field-based ways of doing things and employing people. Another force is the steadily increasing demographic imbalance between developed and developing countries: Many poor nations are seeing population explosions that are generating increased migration, while rich nations can barely replace the populations they have, even as they are beset by job insecurities that make immigration seem threatening. Forces such as these (and others as well, such as the globalization of national and local economies, the spread of terrorism, and the “small” wars that have followed the collapse of the Soviet Union) are bearing down on affluent and deprived regions alike, challenging citizens everywhere to try to understand what is happening; to adapt to a whole new set of changing economic and cultural circumstances; and to do, politically, what appears to be in their own self-interest. In many industrialized nations around the world, there is talk of self-protection, and less attention is being paid both to the sufferings of “outsiders” and to the problem of exclusion within societies. The focus today is much more on differences between people than on the common heritage of the human family.

Second, as developing nations struggle to raise their citizens’ quality of life in the face of these forces, it seems that the broader global questions that should be of great concern to all nations are suffering in consequence. These are the questions of environmental protection; sustainable development; regional or continental cooperation (along with the international legal framework needed for assuring it); the rights of women, children, and the elderly; the treatment of political prisoners; the advent of massive human rights violations; and more. Yet despite their importance, all of these questions appear to be relegated to a much lower priority than are questions directly related, for example, to economic gains to an individual nation or to specific groups within that nation.

In Chapter 14 of this book, Picciotto remarks that “with the end of the Cold War, the developing world has become more volatile and violent conflict more common. Increasingly, [World Bank] resources for development assistance are mortgaged to meet the needs of peacemaking and peacekeeping.” In Chapter 10, Nagao examines from a community perspective “the global economic movement toward an increasingly borderless world, as represented by the opening of national economies to international markets.” He focuses on the social costs of this movement

for rural and urban places, observing, first, the occurrence in rural communities worldwide of a continuing “crisis of identity and intergenerational succession” brought on by the out-migration of these communities’ youth, the aging of their societies, and their sense that in the future, they might “cease to be viable economic entities.” He then points to corresponding effects on urban centers from rural out-migration, especially in developing countries, where sizable flows of people are leading to “excessive population concentration on the receiving end of the cities, causing massive poverty and degradation of the environment.” Similarly, Nieuwenhuysen, in his discussion of immigration policy in Chapter 16, quotes Castles and Miller (1993), who note that “the perspective of the 1990s and the early part of the next century is that migration will continue to grow [and] . . . ecological demographic pressures may force many people to seek refuge outside their own countries” (pp. 3-4).

It is not by accident that world planners and policy makers are substituting sustainable development for the unbounded growth desired in the past. As Zraket and Clark note in Chapter 23:

By the year 2050, the global population is projected to increase by a factor of 2, agricultural consumption by a factor of 4, energy consumption by a factor of 6, and economic turnover by a factor of 8. The cumulative impacts of these estimates in terms of chemical flows are that carbon dioxide emissions will increase by 25%, sulfur emissions by 100%, sediment flows by 300%, water withdrawals by 400%, lead use by 18,000%, and synthetic organic compounds by 70,000%. In terms of ecosystems, fertile soils will decrease by 10%, forested land by 20%, and wetlands by 50%, not to mention unknown impacts upon biodiversity.

Our challenge is to navigate successfully the transition from unbounded growth to sustainability, through research, technology and policy assessment, and critical ex post evaluation of our successes and failures to date.

In all of these areas, evaluation has an important role to play, not only in determining past successes and failures, but in identifying, empirically describing, and monitoring problems; in increasing public awareness of the current and likely future impacts of these problems; and in developing and comparatively assessing proposed solutions. However, it is also the case that evaluation will need to adapt to new conditions, such as greatly modified geographic scope and time periods. For example, assessing

changes in air pollution involves measuring levels and flows across great distances, and evaluating the policy impacts of crop-breeding research on hunger may require as many as 20 years across quite tenuously connected components of the research and technology transfer processes.

The third issue is that, partly because of all these changes and the personal and community insecurities they have engendered, but also because of a late-20th-century social climate that has featured strong grassroots and national movements against taxation in developed countries, citizens everywhere appear intent on contributing less to government and international institutions, and on keeping more of their resources to spend at their own discretion. As a consequence, parsimony in public life and a need for evaluation to justify past expenditures to parliaments, donors, and taxpayers have become common characteristics of the new public management discourse. In Chapter 4 of this volume, for example, Desautels refers to the international “move toward more affordable forms of government” and the “growing public demand for accountability for the results that are being achieved with our tax dollars.” Picciotto also speaks of new “poles of resistance among employers and employees in developed regions who feel vulnerable to the winds of international competition.” “With the end of superpower confrontation,” he writes, “the ideological case for resource transfers to poor countries has evaporated, and the electorates of the industrial countries, preoccupied by fiscal and employment problems of their own, have become far more skeptical and demanding,” calling for “improved performance,” and “documented results” through evaluation. In the same way, Wholey views performance measurement and evaluation as means to rebuild lost public trust in government (see Chapter 8).

In short, the political and policy context in which evaluations are performed is changing. The movement worldwide is toward reduced taxation, reduced deficits, reduced fund transfers from the rich to the poor, and reduced government size. Because of this movement, evaluation takes on greater public importance as the need grows to measure policy and program cost-effectiveness.

But accountability and good management are not the only reasons for conducting evaluations. As the world gets smaller through spreading communications technology and demographic shifts, the opportunity also exists to use evaluation to gain greater understanding and outreach among peoples; better knowledge of the specific problems they confront; more

thoroughgoing dissemination of existing evaluation findings with regard to proposed policy solutions; increased institutional capacity, integrity, and responsiveness; and stronger encouragement of local initiatives—aided by international research, as in Nagao's example of a Japanese town in Chapter 10—in creating innovative solutions to help counteract the new problems of people and communities.

In this sense, evaluation—not only for the purposes of accountability and good management, but also for knowledge building and sharing, for institutional learning and development, for governmental and democratic reform through the serious examination of public policy—has become a precious and unique tool as we prepare to deal with the new socioeconomic, political, and infrastructure needs of the next century. Perhaps it is this more widely glimpsed utility that accounts for the current expansion in many areas now being seen in the evaluation field.

■ *New Actors, New Topics, and New National Evaluation Systems*

Increasing interest by auditors in the use of evaluation for assessing performance led to the 1992 adoption of evaluation as a key work area for the national audit agencies of 188 independent states worldwide, including both developed and developing nations. One important reason for this interest has been the increase in policy demands on auditors everywhere to determine the results of programs—that is, their merit or worth. In consequence, auditors have recognized the need for nonauditing methods that allow the acquisition of information, as Desautels puts it in Chapter 4, “on the more difficult issues (such as attribution) surrounding the outcomes of programs and activities—which is where evaluation has a unique contribution to make.”

In the same way, practitioners of performance measurement (the “new public management,” as it is called in Europe), with its emphasis on assessment and measurement of institutional performance (and citizen satisfaction with that performance), count on evaluation to help achieve governmental or agency reform. In Chapter 9, Mawhood discusses the use being made of evaluation techniques in the United Kingdom to achieve more direct government accountability. For example: