

PUBLIC
POLICY
AND
PROGRAM
EVALUATION

Evert Vedung

PUBLIC
POLICY
AND
PROGRAM
EVALUATION



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PUBLIC
POLICY
AND
PROGRAM
EVALUATION

Evert Vedung

 Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 1997 by Transaction Publishers

Published 2017 by Routledge

2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

711 Third Avenue, New York, NY 10017, USA

Routledge is an imprint of the Taylor & Francis Group, an informa business

Copyright © 1997 by Taylor & Francis.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Notice:

Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Catalog Number: 96-52816

Library of Congress Cataloging-in-Publication Data

Vedung, Evert, 1938-

[Utvärdering i politik och förvaltning. English]

Public policy and program evaluation / Evert Vedung.

p. cm.

Includes bibliographical references and index.

ISBN 0-7658-0687-8 (paper : alk. paper)

1. Political planning.
2. Political planning—Evaluation.
3. Policy sciences. I. Title.

H97.V4313 1997

320'.6—dc21

96-52816

CIP

ISBN 13: 978-0-7658-0687-1 (pbk)

*For Young-seek Choue and Peter Gerlich
without whose invitations
this book would never have come about*



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contents

List of Figures	ix
Preface to the English Edition	xiii
Preface to the First Swedish Edition	xvii
1. Evaluation: A Semantic Magnet	1
2. Evaluation Between Intervention and Feedback	15
3. Evaluation, Rationality, and Theories of Public Management	25
4. Models of Evaluation	35
5. The Eight Problems Approach to Evaluation	93
6. Evaluation Purposes	101
7. Internal or External Evaluation	115
8. Characterizing the Public Intervention	121
9. Monitoring	137
10. Pre-Evaluation	157
11. Impact Assessment as Tryout and Social Experimentation	165
12. Generic, Reflexive, Statistical, and Shadow Controls	195
13. Process Evaluation and Implementation Theory	209
14. Merit Criteria and Performance Standards	247
15. Uses and Utility of Evaluation	265
Glossary of Key Concepts in Evaluation	291
References	313
Index	327



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

List of Figures

Figure 1.1	The Simple System Model	4
Figure 1.2	The System Model Adapted to Government Intervention Evaluation	5
Figure 2.1	Evaluation According to the Spinning Top Model	16
Figure 2.2	Evaluation in the General Governance Model	17
Figure 2.3	Metaevaluation: Evaluation of Evaluation	22
Figure 3.1	Radical Rationalism	26
Figure 3.2	Management by Objectives	29
Figure 4.1	Models of Evaluation	36
Figure 4.2	Goal-Attainment Evaluation	39
Figure 4.3	Side-Effects Evaluation	50
Figure 4.4	Main Effects, Null Effects, Perverse Effects, and Side Effects	54
Figure 4.5	Side-Effects Evaluation with Specified Side Effects	58
Figure 4.6	Goal-Free Evaluation	60
Figure 4.7	Comprehensive Evaluation: The Generic Countenance Model	64
Figure 4.8	Stakeholders in Public Interventions and Their Evaluation	71
Figure 4.9	Productivity	84
Figure 4.10	Efficiency as Cost-Benefit and Cost-Effectiveness	87
Figure 6.1	Dual Loyalties in the Chain of Public Representation	107
Figure 7.1	External and Internal Evaluation	116
Figure 7.2	External Producers of Evaluation	117
Figure 8.1	Policy Instruments: Regulations, Economic Means, Information	123
Figure 8.2	Prohibitions Ordered According to the Strength of the Exceptions	130

x Public Policy and Program Evaluation

Figure 8.3	Some Elementary Types of Economic Tools of Government	132
Figure 9.1	The Fuel Declaration's Envisioned Implementation Chain	140
Figure 9.2	Prerequisites to be Fulfilled if the Fuel Declaration is to Work	146
Figure 11.1	The Impact Problem in Counterfactual Terms	167
Figure 11.2	Evaluation Research Designs for Illumination of Causal Impact	170
Figure 11.3	Classic Experiment with Pre- and Postprogram Measurements	171
Figure 11.4	The Rossi and Freeman Algorithm for Effect Calculation in Randomized Experimentation	173
Figure 11.5	Five Types of Randomized Experimentation	176
Figure 11.6	The Solomon Four-Group Design Applied to an Energy Campaign	178
Figure 11.7	Experiment Without Formal Preprogram Measurement—the Postprogram-Only Control Group Design	180
Figure 11.8	Two Types of Matched Controls	182
Figure 11.9	Theory of the Household Energy Conservation Intervention Field	183
Figure 11.10	Two Quasi Experiments	184
Figure 12.1	Two Reflexive Controls Designs—Interrupted Time-Series and One-Group, Before-and-After—As Amputated Quasi Experiments	199
Figure 12.2	Shadow Controls: Data Sources and Appropriate Questions	206
Figure 13.1	Explanatory Factors in Process Evaluation	212
Figure 13.2	Policy Obscurity	220
Figure 13.3	Motives for Public Policy Obscurity	223
Figure 13.4	An Intervention Theory	225
Figure 14.1	Criteria of Merit for Effectiveness Evaluation	251
Figure 14.2	Four Important Criteria of Merit in Program Evaluation	258
Figure 15.1	The Engineering Model for Utilization of Evaluation	268
Figure 15.2	Learning Theory and Dissonance-Attribution Theory	270

Figure 15.3	Response Steps in Communication Processes	273
Figure 15.4	Strategies to Enhance Utilization of Evaluation	279
Figure 15.5	The Diffusion-Centered Strategy: The Reporting Method	281



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Preface to the English Edition

In August, 1991, I stayed for a couple of nights at the Bildungshaus Sankt Virgil in Salzburg. In the lobby I found a postcard that read in German: "Our life is the history of our encounters." This adage captures important events in the developments that have led to the publication of this book in English.

It all began in 1985 in a workshop on evaluation in Barcelona, arranged by the European Consortium of Political Research, where I presented some ideas from a Swedish manuscript of mine on public policy assessment. In the sessions, I had the exceptional luck of encountering Professor Peter Gerlich, who one year later unexpectedly invited me to lecture at the Vienna University International Summer School, located in the picturesque Alpine village of Strobl am Wolfgangsee, not far from the illustrious Gasthaus Im Weissen Rössl, made famous by Benatzky in the operetta. I decided to teach administrative control and public policy evaluation and have the rest of my unfinished Swedish manuscript translated into English. The 1987 course was repeated in 1989 and 1991.

Before my first visit to Vienna University, however, I had another unexpected opportunity to improve my English manuscript. This happened after the Barcelona meeting in the late fall of 1985, when I taught at the Kyung Hee University's recently inaugurated Graduate Institute of Peace Studies in the Kyonggi Province outside Seoul, Korea. Behind this invitation were Professor Sookon Kim, Professor Kwan-Bong Kim, and Chancellor Young-Seek Choue. The first trip to Korea was followed by two more, in 1988 and 1990, arranged by, among others, Professor Sung-chul Yang. Also, I further expanded and rewrote my English manuscript in a room in the Bursa im. st. Pigionia on Ulica Garbarska, Cracow, during short sojourns between 1987 and 1991, partly supported by the exchange program between

the Uppsala and Jagiellonian universities in the Office of International Affairs, Uppsala, and organized by my good friend and colleague Professor Marian Grzybowski of the Jagiellonian University's Political Science Department.

When I received my fourth invitation from Kyung Hee University to teach in the spring semester of 1994, I decided to finish the manuscript and try to find an English publisher. Since the Swedish book was issued already in 1991, the English version that took form on the flowering Kwangnung campus turned out to be a very different work. The last point should be emphasized: the present work is *not* a mere translation of my 1991 Swedish book, but an entirely revised product.

Through the years, excellent computer literati in Uppsala and Seoul have helped me to install and taught me to use several versions of WordPerfect and PC Tools. In particular I wish to acknowledge the invaluable assistance of Woon-ho Kim, Soon Park Park, Joon-ju Yoon, and Young-chul Kim, all of Kyung Hee University. Dean Yoon-hee Suh, Rector Sung-kuk Lew, and Rector Jae Shik Sohn provided material and moral support. In addition, I take this opportunity to recognize the invaluable help of Pour Mitrahi-Bijän in matters of copying, telefaxing, and mailing, and of my Uppsala colleague Anders Westholm and my competent and spirited brother Tage Vedung in computer matters. The latter has altruistically worked for nights and days to provide the manuscript in camera-ready form.

I have also received inspiration from members of the NOGA Group on research evaluation, headed by Associate Professor Ola Nyquist and financed by the Council for Building Research (BFR) in Stockholm, and from colleagues in the Working Group on Public Policy and Program Evaluation, sponsored by the International Institute of Administrative Studies in Brussels, to which I have belonged since 1992; its chairman and spiritus rector Professor Ray C. Rist of George Washington University recommended me to send the manuscript to Transaction Publishers, advice that I was fortunate enough to heed.

My colleague Alexander Davidson, then in the Uppsala Department of Government, now back in New Zealand, carefully read an earlier version of the whole English manuscript and painstakingly corrected my linguistic inadequacies. Anders Weström did some last-minute reference checking. A substantial part of the work was done when I was the lucky holder of a special research position on evaluation research

in the National Research Council for the Humanities and the Social Sciences (HSFR) in Stockholm. The research has also been supported by the Scientific Advisory Board of the Council for Building Research (BVN) in Stockholm.

To all the people and institutions mentioned, and among the latter particularly the HSFR, I wish to express my profound gratitude. As a special token of appreciation, I dedicate this English-language book to my good friends and respected colleagues Young-seek Choue and Peter Gerlich.

Of course the usual caveat applies: none of the institutions and persons mentioned are in any way responsible for the content of what I have written. The manuscript was finished in January, 1996.

Teatrum Œconomicum,
Uppsala, 10 January 1996



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Preface to the First Swedish Edition

Unintended effects of intentional action fascinate. Few can fail to be gripped by King Pyrrhus' victory or Columbus' voyage to discover the western sea route to India. This phenomenon has been formulated in various ways and observed in diverse fields. Adam Smith thought he saw the Invisible Hand at work, Hegel discovered the Cunning of Reason, and Elster regards friendship and love as essential side effects.

This insight into the reverse consequences and unintended side effects of human planning has been a major force behind the shaping of mechanisms for results-oriented analysis and information feedback in the public sector. Since the 1960s, evaluation has been promoted in democratic states as an important example of such a mechanism. Evaluation has been institutionalized in ministries, public agencies, public investigatory commissions, and particular watchdog agencies.

However, the advance of evaluation has been accompanied by criticisms and crises. Not only have the contents and methodological quality of the evaluations been questioned, but their relevance, inadequate utilization, and even their tendency to "cost more than they taste," if I may use a Swedish proverb. Evaluation researchers have responded by developing new methods and approaches to the application of evaluation in political and administrative decision-making processes. The star of evaluation is currently in the ascendant, not least in Sweden, where so-called results analysis is to constitute the life force in the new triennial budget cycle.

The present work has been written to be used as a qualified introductory volume in evaluation at academic institutions as well as within public administration. It can be used in courses on administration, administrative technique, and organizational management.

My own interest in evaluation, and thereby the production of this

book, is also originally an unintended by-product of the nuclear energy issue.

It was nuclear power that turned me into an energy researcher, which impelled the director of the then Energy Research and Development Commission (*Delegationen för energiforskning*), Sigfrid Wennerberg, to entice me into a project about energy policy instruments. The project, which also put me in contact with Yngve Boye, gradually became directed toward the evaluation of tools of government, and was presented in 1982 in the work *Energipolitiska utvärderingar 1973–1981*.

The continuation was a humble paper on evaluation methods, written in 1982 for the Energy Conservation Commission (*Energisparkommittén*). The inspiration came from the Commission's secretary Ulf Karlsson. The Commission's evaluator, Jan-Eric Furubo, who later transferred to the National Board of Energy (*Statens energiverk*) and who now works at the National Audit Bureau's (*Riksrevisionsverkets*) methods department, has been an indispensable sparring partner in matters of evaluation for more than a decade.

A period at the University of Texas at Austin (1982–1983) helped me to find what has become the American Evaluation Association. Its annual assemblies have not only put me in contact with the ferment of the American evaluation community, but also with a natural catastrophe which could have brought that community to an end, namely the San Francisco earthquake of 1989.

The work has been continued within the research program for public policy "Forskningsprogrammet Offentlig Politik," funded by the Swedish Council for Building Research (BFR) and the Council for Planning and Coordination of Research (FRN), Stockholm. The program director Leif Lewin has not only expressed his views on part of the manuscript, but has also—with unobtrusive efficiency—ensured that his colleagues at the Department of Government in Uppsala have enjoyed a peaceful working environment.

The manuscript has been completed under the exceptional working conditions that a special research position on evaluation research in the National Research Council for the Humanities and the Social Sciences (HSFR) in Stockholm provides.

The Council for Building Research (BFR) and its Scientific Advisory Board of the Swedish Council for Building Research (BVN), through the tandem team Sture Blomgren and Ola Nyquist, have been

a decisive support. No other persons I have come into contact with have worked in such a powerful and goal-conscious way to build up evaluation environments in Sweden. For me, the cooperation with Sture and Ola has been vital. Their evaluation network has provided contacts, infused courage and created material and intellectual conditions for the development of theory and methods. Amongst other members, I would like especially to mention my political science colleague Michael Nydén, who is now working on evaluation questions at the Swedish Agricultural University at Ultuna, Mette Qvortrup, Østfold Distriktshøgskole in Halden, Norway, and guest researcher in Uppsala, and Örjan Eriksson.

In the National Audit Bureau's (RRV's) results analysis community have I found people with similar interests. These people include particularly Rolf Sandahl, with whom I have conducted an ongoing dialogue for several years, and who has read and given comments on the entire final version; Bo Andersson; Ulrika Barklund, who for a few short months was my research assistant; Shirin Ahlbäck, who read the section on the stakeholder model; Anne-Marie Fallenius; and Hans Grohmann.

For a brief period of time the National Council for Cultural Affairs and its work with reviews of national cultural policies in Sweden and other European countries was also a source of contacts and impulses. Valuable insights have been conveyed by Carl Johan Kleberg, Göran Nylöf, and Kjell Eide from Oslo.

The Nordic Summer School of Evaluation Research in Båstad (1985) propelled me into the company of Tore Nilstun and Göran Hermerén. That we engaged Joseph Wholey and Michael Quinn Patton as lecturers opened enduring connections to the American evaluation research elite.

During the passage of years, a great number of friends and colleagues in my own discipline, political science, have supported me with criticism, encouragement, advice, and opinions. Amongst these I would like to mention: Anders Berg; Stefan Björklund; Axel Hadenius; Nils Karlsson; Bo Månson; Olof Petersson; Bo Rothstein and Sven-Erik Svärd, Department of Government, Uppsala; Rune Premfors and Björn Wittrock, Department of Political Science, Stockholm; Ulf Bengtsson, Department of Political Science in Gothenburg; Peter Gerlich, Institut für Politikwissenschaft at the Universität Wien; Hans Ring, Växjö University College; Bo Jonsson and Stig Montin, Örebro

University College; Sven-Olov Larsson and Alf Gunnmo, Östersund University College; and Joachim Schäfer, Kriterium AB in Knivsta. In addition to giving substantive opinions, Anders Westholm at the Department of Government in Uppsala has always been prepared unselfishly to solve troublesome computer problems. In connection to computers, I would also like to name my colleagues Göran Blomberg, Sverker Härd, and Jörgen Hermansson, and last but not least Lars Bruzelius at the Uppsala Data Centre.

Laila Grandin has promptly and reliably sent out conference papers, taken care of travel accounts, ordered reports, and photocopied. Barbro Lewin has maintained order over funding and employees. I suspect that my wife Siv, head librarian at the Psychology Library in Uppsala, has in silence bitterly regretted that she once taught me to borrow and not to buy books, as she has had to locate them in databases and catalogues, and transport them to and from the Carolina Rediviva. That she has furthermore, between rounds of the libraries, bought us a new apartment and given other meaning to my life; I hope that she does not have any regrets.

To all the above institutions and persons I convey my deeply felt thanks. None of those named can be accused of the inadequacies of the presentation, as the responsibility is completely and fully my own.

Uppsala, 12 April 1991

1

Evaluation: A Semantic Magnet

Evaluation:

1. *The action of appraising or valuing (goods, etc.); a calculation or statement of value; = valuation.*
2. *The action of evaluating or determining the value of (a mathematical expression, a physical quantity, etc.) or of estimating the force of (probabilities, evidence, etc.).*

—Oxford English Dictionary, 1933 (1978)

As early as classical antiquity, scholars were summoned to Court to become counsellors to the Prince. Aristotle was hired by King Philip of Macedon as the youthful Alexander's teacher in statecraft. During the siege of Syracuse, the Roman legionnaires were forced to protect themselves from Archimedes' burning mirror and catapults. The tendency has continued in Europe's nation states. Heeding a request from King Christian IV, the prominent astronomer Tycho Brahe took residence at the Court in Copenhagen to read the monarch's horoscope in order to aid the King in the crafting his foreign policy. Members of the Nobel family, in their efforts to invent and supply the Czar with modern weapons, made several tests of explosive substance on the ice of the Neva river in St. Petersburg.

Public sector evaluation is a recent addition to a great chain of attempts by princes to use the brainpower of scholars and scientists to further the interests of the state. The services requested from evaluation experts are, of course, completely different from the ones hinted at in the examples cited above. Evaluation scholars are asked to provide retrospective assessments of the administration, output and out-

2 Public Policy and Program Evaluation

come of government measures in order to effect self-reflection, deeper understanding, and well-grounded decisions on the part of those who are in charge of government operations. Discarding the hackneyed political notion that honorable intentions are enough, evaluation is predicated upon the opposite idea that good practices and solid results are what really count. Evaluation implies looking backward in order to better steer forward. It is a mechanism for monitoring, systematizing, and grading government activities and their results so that public officials in their future-oriented work will be able to act as responsibly, creatively, and efficiently as possible. The interventions of the modern state are so extensive, their execution so complicated, and their potential consequences so far-reaching that science and social research are needed to monitor operations and establish impacts.

However, systematic evaluation is not for contemporary princes alone, but can also be called upon by the political opposition, the professions, the citizenry, or the clientele of government programs. For political scientists, the political opponents' and the citizens' perspectives on evaluation are of particular concern.

Careful retrospective assessment requires systematic data collection, data analysis, and source documentation. In addition, pertinent criteria of merit and standards of performance about how well the intervention must do on these criteria are needed, because evaluation is a normative enterprise.

Evaluation Defined

"Evaluation is the process of determining the merit, worth, and value of things." These words by Scriven (1991:1) capture the basic, natural meaning of the term *evaluation*. Evaluation is the process of distinguishing the worthwhile from the worthless, the precious from the useless.

Evaluation is a key analytical procedure in all disciplined intellectual and practical endeavors. While acknowledging that the process of determining the merit, worth, and value of things permeates every domain of thought and practice, in the present work evaluation will be delimited to suit the demands of public service and governmental affairs. For the purpose of this book, I propose the following definition:

Evaluation = df. careful retrospective assessment of the merit, worth, and value of administration, output, and outcome of government interventions, which is intended to play a role in future, practical action situations.

This definition of evaluation is controversial. Evaluation is circumscribed in numerous ways. Actually, the term evaluation has attracted so many different meanings that we may call it a *semantic magnet* (Lundquist 1976:124). It has come to signify almost any effort at systematic thinking in the public sector. It is easy to agree with the very first sentence in Carol Weiss' early textbook *Evaluation Research* (1972): "Evaluation is an elastic word that stretches to cover judgments of many kinds."

Since evaluation comes in many guises, I shall try to compare in more detail other scholarly definitions of evaluation with the one proposed here. The purpose of the exercise is to put my own definition into a larger perspective. I shall start with the subject matter of evaluation.

Evaluation Concerns Government Interventions

Since evaluation is a truly general analytical process, it can be applied to any area of social endeavor. A special thing in the present context, however, is that evaluation is limited to government interventions only, that is, politically or administratively planned social change, like public policies, public programs, and public services.

Contemporary public interventions cover substantive as well as process-oriented programs (Lundquist 1990). Substantive measures concern diverse functional domains such as energy, environment, natural resources, land use, housing, social welfare, health, transportation, economic development, and many other fields of endeavor. It also includes foreign policy, an area left entirely untouched by systematic evaluation (Vasquez 1986).

Process-oriented interventions—administrative reform—refer to ideas and measures directed at the organization and function of public administration itself (Petersson and Söderlind 1992:7ff). Administrative reform is concerned with management by objectives versus detailed process-oriented management, decentralization, new budgeting

4 Public Policy and Program Evaluation

systems, changes in local administration, and other institution-building processes. A central problem in modern administrative reorganization is which institutional arrangements are used and ought to be used in implementation of public policies and programs: regulatory agencies stacked with neutral and competent executive officials, personnel appointed on political merits, execution through municipalities, corporatist arrangements, professionals, client involvement, or contracting out to private business (overview in Lundquist 1985).

It goes without saying that evaluation embraces the assessment of substantive as well as process-oriented government interventions. Evaluation is targeted at all kinds of public sector activities.

Evaluation is Focused on Administration, Outputs, and Outcomes

As defined here, evaluation is not concerned with the entire policy cycle, but only with the back end of it. To clarify this idea, I shall introduce systems thinking, which is so prevalent in political science study of public administration.

Political scientists tend to view public administration as a system (see figure 1.1). A system is a whole the component parts of which are dependent upon each other. In its most rudimentary form, a system consists of input, conversion, and output in the following fashion:

FIGURE 1.1
The Simple System Model

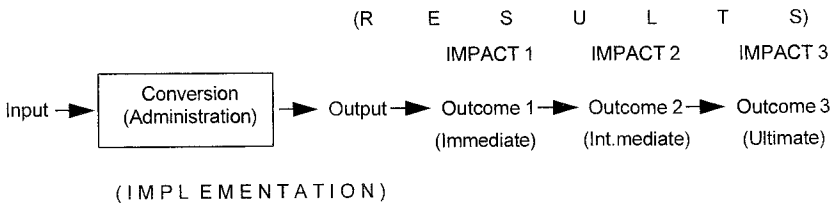


In the field of public administration, the general-systems notion is applied to the civil service, which is viewed as a system. It could be one separate government agency, but also a conglomerate of different organizations. The agencies or conglomerates could be at any level, for instance at the global, the interregional, the national, the intraregional, or the municipal level. The input to an agency from the environment, particularly from its principal (e.g., the government),

may be funds with some strings attached to their use, written instructions, oral support or criticism, and appointed people. Within the agency, funds, people, and instructions are coordinated and converted into something else. The conversion is what is going on in the agency. Output is what comes out of the agency.

In public policy studies, the conversion stage of the general systems model is roughly equivalent to administration, and an outcome phase is often tacked to the output stage of the general system model. By output is meant phenomena that come out of government bodies in the form of, for example, prohibitions, enabling procedures, grants, subsidies, taxes, exhortation, jawboning, moral suasion, services, and goods. Outcomes are what happen when the outputs reach the addressees, the actions of the addressees included, but also what occur beyond the addressees in the chain of influence. We may distinguish between immediate, intermediate, and ultimate outcomes. Another term for outcomes is *impacts*. *Results* will be used as a summarizing term for outputs and outcomes. Results may also indicate either outputs or outcomes. The term *implementation* usually covers conversion and output. The reasoning is summarized in figure 1.2.

FIGURE 1.2
The System Model Adapted to Government Intervention Evaluation



Let me illustrate this with an example. Some years ago the Swedish government instituted a program to help refugees from the civil war in Afghanistan, who lived in camps in Pakistan. To this end, the government allocated funds to the Swedish International Development Agency, abbreviated SIDA. SIDA struck an agreement with the International Red Cross in Geneva, which promised to funnel the money to the National Red Crescent in Pakistan. For the funds, the Red Crescent was instructed to buy tents and blankets from local dealers, and provide the equipment to the refugee camps. In the camps, the local Red Crescent branch was expected to put up the tents, and distribute the blankets to the refugees. Then, refugees were supposed to use the tents and blankets in order to alleviate their plight.

To qualify as an evaluation, a study of the Help-to-Afghani-Refugees Program must concentrate on either the outcome (if the refugees were actually using the tents and the blankets and if that alleviated their plight), the output (the distribution of tents and blankets through the local Red Crescent), or the administration (what happened to the funds once they had reached the SIDA through the purchase of blankets and tents by Pakistani authorities). Admittedly, outcome evaluation may be considered more important than output or conversion evaluation. However, I do not want to equate evaluation with outcome evaluation. The concept, as defined here, includes concern with administrative processes and output as well. In the refugee case, for instance, everybody can see that administration is a long process with several levels of authority involved.

The limitation to outcomes, outputs, and administration excludes studies assessing *ex post* the merits and drawbacks of features in the policy formation phase. For instance, actual or past policy formation can be assessed against such evaluative criteria of merit as comprehensive and reliable information base or participation by various affected interests. One may evaluate measures on the books, using such dimensions of merit as comprehensibility or consistency with other programs. In this context, however, such studies will not be reflected upon.

One additional clarification is probably justified. The limitation to administrative processes, outputs, and outcomes is not concerned with explanatory factors in evaluation. If the evaluation sets out to explain what influenced variations in administrative procedures, outputs, and outcomes, my definition allows for explanatory factors to be drawn

from anywhere. It would be abjectly inappropriate to delimit the concept of evaluation with respect to the determinants that may be discerned.

Now, I have ventured to justify the delimitation of the subject matter of evaluation to “administration, outputs, and outcomes of government interventions.” In the next section, I shall address what it means for evaluation to be “retrospective”.

Evaluation is Retrospective

Evaluation is retrospective assessment of public interventions. Prospective appraisals (i.e., scrutinies of courses of action considered but not yet adopted even as prototypes), are not included in my definition. Also this limitation is controversial, particular in the North American context. Leading theorists argue that prospective assessment—*ex ante* assessment, forethought evaluation, needs assessment, analysis for goal-setting—does belong to evaluation. To them, evaluation becomes an umbrella, covering all kinds of analyses of, in, and for public intervention. Is it reasonable to let “evaluation” refer to almost any intellectual effort in the public sector? Cases of this large perspective on evaluation can be taken particularly from economists, who perform cost-benefit and cost-effectiveness analyses of potential, future options, maintaining they practice evaluation. Also Rossi and Freeman (1989:18) adopt this large perspective when they maintain: “Evaluation research is the systematic application of social research procedures for assessing *the conceptualization, design, implementation, and utility of social intervention programs*” (italics mine; also Anderson and Ball 1978:3, 11, 15ff).

“If planning is everything, maybe it’s nothing?” Aaron Wildavsky (1973) ironically wondered two decades ago on the then strongly fashionable fad, planning. Can the same question be raised today, when *ex ante* assessment is included in evaluation? Of course, it is both futile and foolish to legislate about the use of a word. But if evaluation is allowed to embrace all kinds of analysis in political and administrative life, will not the concept become too diluted? Here, perhaps, we face another instance of the semantic magnetism of the word *evaluation*.

The major argument against including *ex-ante* assessments in evaluation is drawn from the origin and history of evaluation research. The demands of the early evaluation movement for empirical data on policy

and program results emerged in opposition to the prevailing emphasis on analysis of planned interventions. If evaluation is allowed to embrace even planning, this significant historical line of conflict will be obscured.

Hence, in this context I have confined evaluation to after-the-fact assessments. Such assessment concerns adopted interventions in the sense of ongoing or terminated policies, programs, program ingredients, and the like, no matter whether they are veterans or recently introduced small-scale prototypes. Before-the-fact analysis of potential, not-yet-adopted interventions, however, is not included in evaluation in my usage in the present book.

Evaluation is Assessment of Ongoing and Finished Activities

Sometimes evaluation is restricted to *ongoing* activities, leaving out assessment of finished policies and programs. This quite narrow perspective is clearly discernible in David Nachmias's textbook *Public Policy Evaluation* (1979:3f):

One method that can reduce the number of erroneous decisions is the formal scientific approach to knowledge. . . . Viewed from the scientific perspective, policy evaluation research is the objective, systematic, empirical examination of the effects ongoing policies and public programs have on their targets in terms of the goals they are meant to achieve.

Actually, the narrow perspective on the subject matter of evaluation is a commonplace in American and Canadian literature. According to Rutman, "program evaluation refers to the use of research methods to measure the effectiveness of operating programs" (1980:17). Wholey et al. write: "Evaluation assesses the effectiveness of an ongoing program in achieving its objectives" (1970).

Indubitably, ongoing interventions clearly belong to the subject matter of evaluation. They may even constitute the core subject of public sector evaluation. But should evaluation be confined to ongoing activities *only*? The answer must be no. Lessons can be learned also from assessments of terminated operations. Why exclude these from evaluation research? I would argue that careful assessment of terminated public activities—summative evaluation to use Michael Scriven's famous expression—may very well assist in improving ongoing operations. Terminated policies and programs may also be scrutinized

for accountability. Delimiting evaluation to the study of ongoing activities would unduly fence off important parts of policy evaluation research. Evaluation should include all retrospective or *ex-post* study of policies and programs, ongoing as well as terminated. According to my definition, evaluation comprises all kinds of *ex-post* analyses of administrative processes, outputs, and outcomes of public sector activities.

Evaluation is More than Impact Assessment

On some rare occasions, eminent methodologists have also defined evaluation as impact assessment. In his *Planning Useful Evaluations*, Leonard Rutman (1980:17) states:

In this book program evaluation refers to the use of research methods to measure the effectiveness of operating programs.

And effectiveness, he adds, is the extent to which a program achieves its goals or spawns certain effects.

Thus, evaluation is restricted to studies raising the impact issue. I do not accept this unduly narrow definition here. In my view, apart from impact assessment, evaluation should also include efficiency assessment, administration and output monitoring, as well as simple goal-achievement measurement where the impact issue is not raised. Furthermore, in the evaluation of research, libraries, museums, public health, public parks, and other government services, the quality of the output is assessed, for instance, by expert panels, peer arrangements, or client groups, but in these studies the causality issue is frequently not raised at all. Still, the studies are referred to as evaluation, and justifiably so. Eventual criticism of these practices ought to be advanced by means of substantive argument, not by definitional fiat.

Intervention Goals are not the Only Permissible Value Criteria

In the discourse on evaluation, there isn't even general agreement on a seemingly fundamental point: that evaluation is concerned with the determination of merit, worth, and value. Charles Atkin, for instance, includes no reference to valuing in his definition, which goes like this: "evaluation is the collection and presentation of data summaries for decision-makers" (quoted from Scriven 1991:156).

My position is that evaluation by definitional fiat involves the assessment of merit, worth, and value. The value component of evaluation presupposes at least one criterion of merit against which public interventions are judged. Some authors, particularly in the past, took a rigid view of what value criteria were permissible for something to be called an evaluation. Evaluation was equated with assessments against premeditated, avowed *intervention goals*. Argues Carol Weiss (1972a:4):

The purpose of evaluation research is to measure the effects of a program against the goals it set out to accomplish as a means of contributing to subsequent decision making about the program and improve future programming. (See also quotation from Nachmias above.)

Admittedly, the traditional formulation of the program evaluation question is: To what extent does the program succeed in reaching its goals? But why should intervention goals be the *only* permissible criteria in judging the merit of public policies? Why shouldn't assessments using other criteria of merit than intervention goals be referred to as evaluation? Why shouldn't judgments using concerns and expectations of the clients or of other stakeholding audiences be called evaluations? Why shouldn't appraisals using yardsticks drawn from political philosophy—that is, equality, freedom, and justice—be regarded as evaluations? The selection of criteria of merit is a crucial, albeit little debated, issue in evaluation. Excluding all yardsticks except the avowed intervention goals is in my opinion to circumscribe the concept too narrowly. It will by fiat preclude discussions concerning what criteria are to be applied to our evaluands. Therefore, the definition propounded here contends that evaluation engenders appraisal but leaves open what kinds of value criteria may be used.

I wish to make another point with regard to criteria of merit. Evaluation may apply either descriptive or prescriptive theories of valuing, to use the illuminating distinction by Shadish, Cook, and Leviton in their masterful treatise *Foundations of Program Evaluation* (1991:46ff). A descriptive theory of valuing implies that evaluators use the merit criteria of others—such as intervention goals—as evaluative yardsticks, as opposed to a prescriptive theory that advocates the primacy of particular values. My definition is neutral with respect to these two approaches. It is also neutral in relation to subdivisions within the two approaches. Within descriptive valuing, some authors prefer to use

premeditated goals, whereas others suggest a kind of division of labor on this account between formal evaluators and recipients of the formal evaluations. In complicated cases, formal evaluators may resist from applying premeditated value criteria in the expectation that the assignment of weights to the information provided is best performed by the recipients afterwards once the study is finished. Also this would pass as evaluations according to my definition. In order for something to be an evaluation, it is not necessary that the researcher-cum-evaluator actually performs the valuing. Some valuing must be executed, but it might be executed afterwards by some recipient. If this is what Atkin had in mind, the difference between my position and his is probably negligible.

Evaluation is Careful Assessment

Evaluation is *careful* assessment of public interventions. This means that an evaluation must meet some minimum standards of quality such as systematic data collection and the conscientious application of criteria of merit and standards of performance.

Many celebrated authors have proceeded further and made the explicit point that evaluation by definition is social research. The following statement by Elinor Chelimsky (1985:7) is typical for these scholars:

Program evaluation is the application of systematic research methods to the assessment of program design, implementation, and effectiveness.

It is obvious that Chelimsky is preoccupied with cases where the canons of social science methodology are used to make the judging process more accurate. She attempts to draw a sharp line between evaluation research and offhand evaluations that rely on intuition, opinion, and trained sensitivity. The same attitude is displayed by other authors as well. In his pioneering *Evaluative Research* (1967:12; cf. Mann 1972, Caro 1971), Edward Suchman stated: "We do not view the field of evaluation as having any methodology different from the scientific method; evaluative research is, first and foremost, *research* and as such must adhere as closely as possible to currently accepted standards of research methodology" (also Meyers 1981:50ff; Nachmias 1979:1ff).

I disagree with Chelimsky, Suchman, and others that evaluation

should be minimally *defined* as research, although I use research in a relatively wide sense. Research is not equated to science and the application of quantitative methods. It also includes humanistic research and qualitative methods. Furthermore, research encompasses commissioned research as well as research under the scholars' own program responsibility. Nonetheless, I shall not equate evaluation with research, because there are alternative ways of regarding evaluation which it would be outrageous to discard by linguistic stipulation only. For example, evaluation is often an integral part of the political and administrative decision-making processes in which careful research plays a role, albeit not a dominant one. I shall also refer this to evaluation, provided there is some careful *ex post* assessment involved.

Evaluation Should be Useful

On a general level, evaluation should be delimited according to purpose as well. Evaluation is not any careful appraisal; it is appraisal intended to play a role in future practical action situations. By this I do not necessarily mean that a study must have immediate relevance to a pending decision to be called an evaluation. The purpose may be to contribute to public policy debate, that may or may not lead to some authoritative decisions. A straightforward aim, on the other hand, is to perform evaluation to provide reasoned judgments based on solid empirical ground to be employed in decision making. Another, more dubious intention, would be to conduct evaluation to give outsiders the impression that the organization is handling things well. The general practical orientation is so central that it must be included in the definition.

On the other hand, further specification of the practical orientation would make evaluation too circumscribed. Some veterans of evaluation research argue, for instance, that evaluation always aims at intervention refinement. Lee Cronbach et al. (1985:14), among others, have defined evaluation with respect to the intervention improvement idea:

By the term *evaluation*, we mean systematic examination of events occurring in and consequent on a contemporary program—an examination conducted to assist in improving this program and other programs having the same general purpose.

True, future intervention betterment is an important purpose of evaluation. The problem with Cronbach's definition is that numerous evalu-

ation studies are undertaken for practical purposes other than improvement; still, it seems reasonable to call them evaluation. Exercising evaluation for accountability reasons, that is, principals undertaking evaluation with the aim of holding their executives responsible for what they have done, is a perfectly valid purpose for evaluation (see chapter 6). True, accountability studies can lead to intervention improvement. However, they can also end up with intervention termination. Furthermore, evaluations are undertaken for political purposes, for instance, to save time or divert public interest from the issue. In order not to preclude far-ranging and deep-probing discussions on why evaluation should be exercised, it seems preferable not to limit evaluation only to the intervention improvement purpose. By definition, evaluation is an enterprise intended to play a role in future practical action situations, such as, for instance, kindle public policy debate or more directly provide useful materials for some pending decision. However, the concept ought not to be further delimited with regard to purpose.

This ends my justification for defining evaluation as careful, retrospective assessment of merit, worth, and value of the administration, output and outcome of government interventions, which is intended to play a role in future practical action situations.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>