Extraits !

# PUBLIC POLICY AND PROGRAM EVALUATION

*Evert Vedung*

TRANSACTION PUBLISHERS
New Brunswick (U.S.A.) and London (U.K.)

Second paperback printing 2004


This book is printed on acid-free paper that meets the American National Standard for Permanence of Paper for Printed Library Materials.

Library of Congress Catalog Number: 96-52816
ISBN: 1-56000-299-9 (cloth); 0-7658-0687-8 (paper)
Printed in Canada

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

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

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-

# Contents

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

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**

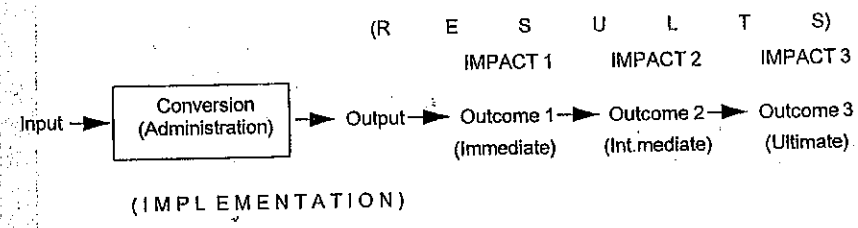

In the field of public administration, the general–systems notion is applied to the civil service, which is viewed as a system. It could be one separate government agency, but also a conglomerate of different organizations. The agencies or conglomerates could be at any level, for instance at the global, 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 theoretists 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 evaluation 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.

# 2

# Evaluation Between Intervention and Feedback

Evaluating engenders looking backwards to improve forward direction. The specific role of evaluation is to systematically amass and assess information on intervention outcomes, outputs, and administration to produce adjustments, or more rational future decisions.

## Spinning Top Models

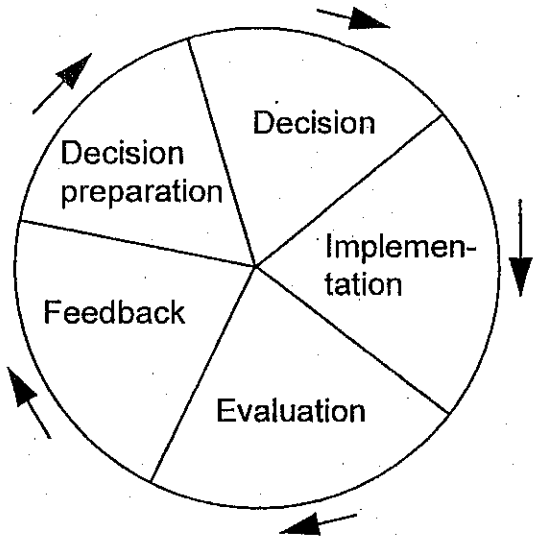
To illustrate this ex post facto orientation, the public intervention process is usually regarded as a cyclical operation divided into a number of stages, where evaluation constitutes or is included in a later such stage, which points to the next decision cycle. The stage models are often drawn as spinning tops, cycles or spirals. (see Figure 2.1) Presumably, the stage models do not purport to picture public sector realities, but are heuristic devices to facilitate thinking or tools by which realities can be investigated.

A drawback with the spinning top models is that public policies and other public sector activities seem to swirl around on the same spot. There is no development or direction involved. Another demerit is that evaluation is on apar with all other processes.

## Intervention Process, Evaluation Process, and Feedback Process

A superior way of illustrating the afterthought idea is to detach the evaluation process from the intervention process and view them as two different subprocesses of the larger political– administrative gov-

FIGURE 2.1
Evaluation According to the Spinning Top Model

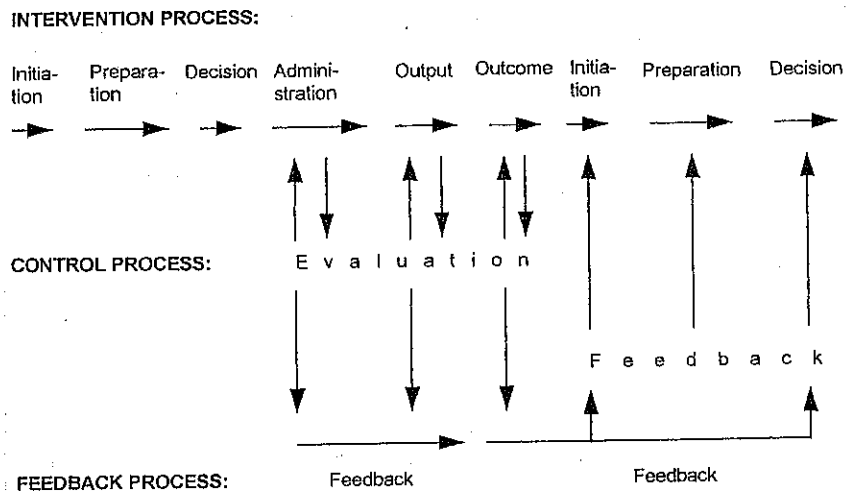



ernance system. The evaluation process in turn can be regarded as one component of a more general administrative retrospective control system. Administrative control is the means used by an organization to elicit the performance it needs and to check whether the quantities and qualities of such performance are in accord with organizational specifications. Evaluation, as part of such a control system, is a systematic reflection on the final stages in the primary public intervention process. One of the crucial problems in governance processes is to link this secondary process—metaprocess—to the primary intervention process. This is achieved through a feedback mechanism in which the evaluation is utilized. Also the feedback can be perceived as a separate process. In the governance model presented in figure 2.2, the primary public intervention process stages are limited to six, with evaluation and feedback drawn as two distinct metaprocesses.

It must be stressed that the governance model does not provide an empirical description of any reality. It is a heuristic, an instrument to support thinking. It should guide reflection and research on public intervention processes, evaluation and utilization of evaluation. It is a tool of analysis designed to provide hypothetical insights, and create ideas on appropriate research problems. The model neither contributes a realistic picture of public policy or evaluation reality, nor is this its purpose.



FIGURE 2.2
Evaluation in the General Governance Model




In general systems terminology, accounted for in the previous chapter, initiation, preparation, and decision in the intervention process would belong to the input stage; administration would be conversion and output would be output. Outcome is usually not covered in the general systems model, but must be included in any heuristic which purports to be fruitful for evaluation purposes.

According to the general governance model, the public intervention process starts with *initiation*. Initiation concerns the identification of a problem, the solution of which supposedly demands public action. Demands for government consideration may arise from within the political system, by for instance an administrative body, or from some external source such as interest organizations or the media. When political officials address the problem it has entered the political agenda. When national politicians start to become concerned and attempt to

bring it up in some national political channel, then it has been brought up on the national political agenda.

Aside from the substantive issue, *decision preparation* is concerned with implementability of proposed measures and evaluability of results to be achieved. First, efforts are made to pinpoint the nature of the problem for the purpose of creating a base for an upcoming decision. Information concerning alternative courses of action is gathered and organized, the presumed consequences of the alternatives are calculated, the costs of the consequences are assessed, goals are set, and the alternatives are assessed in the light of what is to be achieved. Second, decision preparation is attentive to before–the–fact implementability in the sense that investigations are performed to determine the legality of the options, and the available organizational, managerial, and manpower capabilities in order to establish the degree to which the proposed intervention alternative can be specified and implemented. Third, decision preparation may also involve considerations concerning how a future intervention should be evaluated. Ideal–type decision preparation, then, not only includes substantive concerns with goals, means, costs, and benefits, but also amenability to successful implementation and evaluation.

Decision preparation may involve numerous actors and procedures for example appointment of a formal investigatory body, presentation of preliminary options for action, review by interest groups and other stakeholders, reworking of the original proposals, presentation of one or several final suggestions for action. Decision preparation ends with the presentation of proposals for decision.

Decision preparation, consequently, is concerned with goals, means, costs, benefits, implementability, and evaluability. But it also involves actors, conflicts, clashes between different stands of the issue, negotiations, and compromises. Preparatory work for political and administrative decisions can never be turned into pure intellectual problem solving. Invariably, it also contains features of conflict and conflict resolution.

*Decision* implies that a formal, authoritative, legitimating resolution is made. Who will make this resolution varies of course from intervention to intervention. If the decision concerns legislation is must be made by the parliament. Furthermore, intervention decisions can be made by the government, the national agency and the regional agency. At the municipal level, decisions are taken by the municipal councils,



the municipal commissions, or the municipal agencies. The decision function is much less comprehensive than for instance the decision preparation stage.[1]

Decision in my heuristic model concerns the formal intervention decision. Where the real decisions about the adoption are made is held open for empirical investigation. Often the real decision is made already in the preparation stage.

*Administration* is a wide and diversified category. The basic idea is that intervention decisions should be brought forward to their realization. Even though decisions to intervene and administrative decisions to achieve the purposes of intervention decisions are sometimes hard to differentiate, it is important to make the distinction. Through administrative decisions, efforts are made to plan, design and make interventions ready for delivery to the targets. This can be thought of as the core process of administration. The core process needs support processes, however, like for instance hiring and training of personnel, purchase of technical equipment, payments of salaries, and gatekeeping functions like communication with the outside world.

Administration involves intervention specification and preparation for field execution. Specification includes such things as diffusion by upper–level decision makers of information on the contents of general decisions, specification by governments of general regulatory regimes adopted by parliament, and specification by regulatory agencies of regulatory mandates through the issuance of detailed norms and rules. Preparation for field execution refers to political attempts to affect government agencies, upper management attempts to influence lower management, and lower managements attempts to influence street–level operators. The addressees may also contribute to administration, for example, through active participation in the formation of the grass–roots agency's intervention delivery. In public administration language, this is called client or user participation.

Outputs are the means through which street–level operators and other agents in the administrative system attempts to influence intervention targets. Outputs are the things actually pouring out of the administrative system. Outputs include social services and goods delivered, laws and ordinances applied to individual cases, taxes and levies collected, grants and loans disbursed, campaign messages distributed to targets, and the like. In more concrete terms, what happens is that teachers teach in the schools, senior citizens are taken care of in

the nursery homes, highways are built, trash is hauled, and books are borrowed from the public libraries.

Sometimes, outputs are included in the administration stage. The reason why they are treated separately here is a wish to facilitate for the readers to see the parallel between the public intervention model and the general, simple system model. Since the simple system model clearly differentiates between conversion and output (see figures 1.1 and 1.2 above), I have also made the separation in my heuristic intervention model.

*Outcomes* are what happens on the addressee (client, recipient) side in the intervention model. Outcomes can be measured in several stages. In some circumstances, they are gauged as addressee attitudes, addressee actions, and addressee urges to act. In other situations outcomes are essentially constituted by primary, secondary, or even tertiary societal repercussions beyond the addressees. Outcomes are then thought to be effects of addressee actions.

It may be difficult to pinpoint exactly the factors in the governance process that in a special case actually produce an outcome. An outcome may ensue because some addressees anticipate a future intervention decision and adjust before they are formally obligated to do it. The adopted intervention may directly impact upon the addressees, as when they without compliance measures comply with a regulation. Finally, the outputs may also produce the outcomes in the general population. What is important to realize is that outcomes are not always produced only through formal administration of higher level policies and outputs, but also through anticipation before the formal intervention decision is made.

In all organizations, a fundamental management task is *after–the–fact control*. Some device for monitoring and redirecting the diverse and specialized activities of large complex organizations such as a public bureaucracy is necessary, if the system is to be effective. In the present context, after–the–fact control is viewed as a metaprocess, as a separate, ongoing cycle of monitoring the primary public intervention process, including the collection of information about the performance of subordinates, and the screening of addressee compliance. After–the–fact administrative control comprises traditional auditing, simple monitoring, and evaluation. Administrative control also includes self–control in the sense that intervention operators themselves may control their work and work results in order to be able to adjust and accommodate.[2]



*Evaluation* is one type of after–the–fact administrative control. Decision makers, upper management, lower management, and street–level operators authorize their subordinates or external researchers and consultants to find out how the various stages in the administration process are unfolding, what the outputs and outcomes look like, whether the outcomes are produced by the intervention, and whether there are more cost–efficient means to reach the same goal. Evaluation may also be conducted as self–assessment for the purpose of learning and self–adjustment.

Findings dug out through after–the–fact control and particularly evaluation are supposed to be used as information for learning in the primary intervention process. Evaluation findings might be communicated to different audiences, such as agency managers, program operators, or outside decision makers. By itself a metaprocess, the delivery of control information is called *feedback* in figure 2.2. On the basis of the information supplied, the various evaluation users can take action in or directed at the intervention process. Feedback is supposed to produce redirection and reconsideration. In this stage of redirection and reconsideration, three things can happen as a consequence of learning: continuation as before, change, or termination. On the basis of these new circumstances, the intervention process is supposed to run another round and be subjected to after–the–fact control, which gives rise to new feedback.

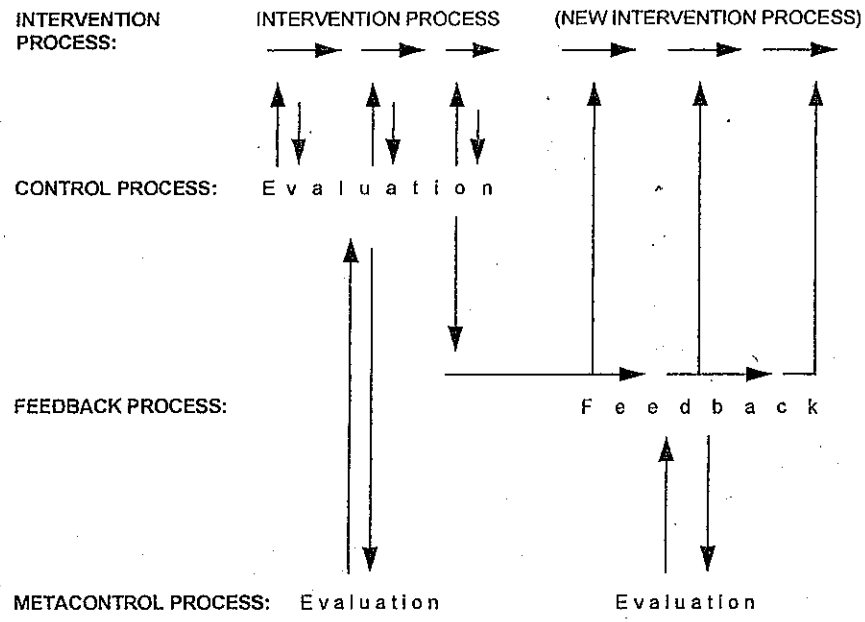
## Metaevaluation

As a final note, I would like to add that metaprocesses like evaluation and feedback can be evaluated as well. At a first glance, auditing evaluation may seem a peculiar activity. But evaluation, as a part of management in an organization, is as auditable as other management processes, such as policy implementation. To continue with the terminology previously used in this chapter, metaevaluation is a metameta–process in the general governance process. The relationships between the various levels are illustrated in figure 2.3.

Also, insights produced by evaluation of evaluation and feedback evaluation should be fed back to appropriate decision makers in order to be utilized. This metafeedback mechanism is not illustrated in figure 2.3.

Metaevaluation may be applied to one particular evaluation, for

FIGURE 2.3
Metaevaluation: Evaluation of Evaluation




example, before it is sent to the commissioner or prior to its publication. It may be carried out by the evaluator herself, by some external scrutinizers, or preferably by the evaluatees. The purpose is to check for methodological quality, readability, faithfulness to facts, and other properties. Metaevaluation may also be carried out as a summary of findings from several evaluation studies.

Still another possibility is to audit the evaluation function of some government agency. A frequent subject in the evaluation literature (Hoogerwerf 1992:215ff), metaevaluation in this sense is especially appropriate when superiors want to evaluate the performance of some subordinate body with strong professions. University research is a perfect case of such a government supported activity where professionals reign almost supreme. The idea is then to let the professionals themselves manage the evaluation function—professors are asked to



evaluate other professors and their performance, because outsiders do not have the necessary competence—but that the superior decision maker may ascertain whether the evaluation mechanisms actually function and function well.

Let me once more emphasize that the governance model is a heuristic model, which is supposed to facilitate thinking on evaluation. It is supposed to help evaluators and commissioners of evaluation to place themselves into a larger context. It neither provides not does it purport to provide a realistic picture of what is going on in governments. The realistic picture can be painted only on the basis of careful empirical studies.

## Notes

1. Decision stands for program decision. Evidently numerous decisions are made also in the other stages in the primary governance process. Decisions are made to appoint investigatory commissions, how implementation should be carried out, how ex post facto control should be organized, and so on. This obvious complication will be overlooked here.
2. The idea of regarding evaluation as a meta–process is clearly expounded in Fernández–Ballesteros (1992a:207), who distinguishes the public policy intervention cycle from the evaluation (assessment) functions. Also in the same author's 1992b.

# 3

# Evaluation, Rationality, and Theories of Public Management

All evaluation rests upon the minimal rational supposition that goals, intentions, perceptions, judgments, opinions, schemes, plans, even ideologies—in short everything concerned with the world of human consciousness—play such interesting roles in public sector action, that their functions are worth investigating. Humans are seen as calculating actors, capable of choosing among options and judging, albeit crudely, their consequences. Through experience they may learn from past actions.

In recent times, attempts have been made to marry evaluation to various top–down management doctrines, particularly management by objectives and management by results. This is by no means necessary. Evaluation can be part and parcel of more bottom–up or pluralist management doctrines as well, such as professional evaluation, client–oriented evaluation, and stakeholder evaluation (see chapter 4). In addition, evaluation may be a one–shot event. In this chapter, I shall scrutinize briefly the top–down management doctrines, and discuss other possibilities in the next chapter.

## Evaluation and Radical Rationalism

Elementary evaluation is ancient. According to Daniel's account of the Babylonian Imprisonment, King Nebuchadnezzar's courtier, Aspenas, systematically tested if a special diet given to the Jews Daniel, Hananya, Misael, and Asarya had any effect. As controls, Aspenas

used Babylonian adolescents, who were given the King's food to taste (Dan.1:1–21; and Chelimsky 1985:2)

Yet, however ancient after–the–fact assessment is, contemporary evaluation research undoubtedly was part of a mighty rationalistic tide, the main philosophical thrust of which was to convert politics into a more rational, even scientific enterprise. This rationalistic, technocratic tendency grew particularly strong in the 1960s and the first part of the 1970s. According to dominant wisdom between 1960 and 1975, makers and administrators of public policy ought to apply the whole spectrum of methods for program–budgeting, zero–based budgeting, strategic planning, futures studies, systems analysis, cost–benefit and cost– effectiveness analysis, methods commonly referred to as policy analysis. Then ever–changing, fickle, and ill–considered public policies might be avoided and government interventions would become well–grounded, sound, and efficient. Through use of science or sciencelike analysis, well–co
nceived, long–range, and reasonable public policies would influence, if not replace, the shortsighted games between parties and special interest groups, which through simple rules of thumb attempt to muddle through arising societal problems (Wittrock and Lindström 1984).

The most far–reaching element of this rationalistic current in politics and administration was "radical rationalism" (Wittrock and Lindström 1984). Radical rationalists argued that decisions should be made only once researchlike policy analysis had answered a series of questions, somewhat simplistically summarized in figure 3.1.

**FIGURE 3.1**
**Radical Rationalism**

1. What ends are the decision makers trying to reach and what is the problem in need of a solution?
2. What alternative options may contribute toward achieving the end?
3. What are the consequences of the different options and the probabilities of each of these consequences?
4. What are the costs and resource requirements of the various options?
5. How can the options be arranged with respect to costs and consequences, and which criterion of merit should be used in the choice of option?



In the United States, Wittrock and Lindström (1984:11ff) continue, radical rationalism originated at the Rand Corporation and spread via Robert McNamara's Department of Defense to units of planning and analysis throughout the Federal administration in Washington. In Sweden, the diffusion was accomplished by systems analysts in the National Defence Research Establishment (FOA), the Agency for Administrative Development, and the National Audit Bureau. In both cases, the defense research communities seem to have played a role. Streamlined planning systems were created for physical, regional, and economic planning, for traffic and environment, higher education, and energy. Thinking in terms of clear goals, systematic data collection on problems and exhaustive before–the–fact analysis of measures dominated.

The combination of researchlike analysis and extensive programs marks the breakthrough of what Wittrock and Lindström call "the era of the great programs." While in the United States the new methods were part of president Lyndon Johnson's Great Society, in Sweden, they were intended to strengthen the Social Democracy's "Strong Society."

Radical rationalism, in my interpretation, emphasized the future–oriented, decision–preparing, planning stages in public decision–making operations. The essential requisite was that interventions were wisely and synoptically designed in a central planning machinery before they were established; execution and actual goal achievement were seen as relatively unproblematic. Actually, radical rationalism's trusting belief that results come up as effects of human design bears strong resemblance to what Popper and Hayek has called "naive rationalism" as opposed to "critical rationalism," which accommodates to the occurrence of unexpected side effects, reverse effects, and other limits to omniscient rationality.

With time, trust in radical rationalism's central–planning euphoria eroded. Observers realized that beautifully crafted plans are one thing, their transformation into practical reality another. More and more, administrative pundits started to argue the necessity of before–the–fact implementation analysis as a necessary missing link of rationalist policy analysis. And experts on administration came to the conclusion that a retrospective evaluation function had to be established in public policy, since reforms could lead to unexpected null effects, perverse effects, and side effects, while costs were skyrocketing.[1]

## Evaluation and Management by Objectives

It was partly in this context that the focus of interest shifted toward *management by objectives,* and away from older notions like input–oriented and process–oriented management. Management by objectives incorporates three features that are thought to constitute good management practice in government: setting of clear goals, participation in decision making, and objective feedback of achieved results.

Management by objectives not only asserts that parliament, government, top managers, and other principals should set clear, measurable, intervention objectives. The objectives should be substantive and refer to results. Second, top management should involve senior administrators and their subordinates all the way down to the lowest hierarchical levels in objectives–setting discussions, which will result in the development of successively more specific objectives at all levels. Preferably the objectives–setting discussions should have a give–and–take character. Finally, efforts should be made to monitor progress toward the objectives and evaluate the results in terms of effectiveness. These results should be fed back to top management and senior administrators, but also to low–level subordinates because in the management–by–objectives doctrine, the latter are supposed to react to evaluation results and adjust their behavior accordingly. In the Swedish version of management by objectives, currently under implementation, a fourth step is added: those levels that are most successful in fulfilling the objectives should be awarded; the unsuccessful ones punished.

The stages of a management by objectives process are outlined in figure 3.2 (adapted from Rosenbloom 1989:159f).

The opposite to management by objectives is process–oriented management. In *process–oriented management,* politicians formulate rules for the decision–making processes in the public agencies and other executive bodies in the hope of influencing output, outcome, or costs. Immediate objectives are targeted at internal administrative processes, not at results.

While placing strong emphasis on participatory goal determination, management by objectives on the other hand attaches less weight to the setting of goals for the input into or the decision–making processes within the public agencies; instead it emphasizes what should come out in the form of impacts on society and to some extent outputs. Another crucial feature is learning from feedback and knowledge of



**FIGURE 3.2**
**Management by Objectives**

1. The top managerial level determines overall, organization–wide goals and sets priorities and posteriorities among the goals; the goals are formulated in terms of results to be accomplished, preferably as outcomes but also as outputs;

2. Top management principals involve lower–level agents—such as senior administrators, middle managers, and their subordinates—in the work of breaking down the overall organization–wide goals into subgoals for each organizational unit, even for each individual, operationalizing the subgoals into fully measurable results, making priorities and posteriorities between the subgoals, and indicating the time frame within which the subgoals are to be achieved; top managers are actively supporting and participating in the objectives–discussion;

3. Units and individual staff members develop plans for accomplishment of the intended results;

4. The superior liberates his subordinates from directives concerning how the goals ought to be achieved so that the subordinates themselves can decide by which means and methods the subgoals and objectives should be accomplished;

5. The principal allocates resources to his agents, preferably as block grants—unspecified lump sums—to provide some degree of freedom in developing means for the achievement of objectives;

6. People in the agencies are involved in implementation of plans, with emphasis on communications for responsiveness and on broad sharing in establishing authoritative objectives, priorities and posteriorities;

7. A system of performance review is instituted to track progress toward goals and objectives, with specific intermediate milestones indicated;

8. The results are frequently reviewed and evaluated in terms of effectiveness;

9. Evaluation findings are disseminated not only to top–level management, senior administrators, and middle managers but all the way down to the staff members at the lowest hierarchical levels of the organization to enable everyone to compare his performance with others in order to generate and implement improvements in objectives and results;

10. Successful work toward goal accomplishment is awarded by the principal (salaries, merit pay, promotion), shortcomings and failures are punished.

results. It is thought that both goal setting and performance evaluation will contribute to increased productivity. Some researchers have actually found that "it is the goal originally established that produces most of the 'motivational force' in the situation rather than the provision of feedback alone" (Carroll and Tosi 1973:4).

## Evaluation and Results–Oriented Management

Evaluation can also be seen as a vital feature of *results–oriented management* (Wholey 1983; Shadish, Cook, and Leviton 1991:233). Like management by objectives, the term results– oriented management is ambiguous. I have to admit that I am not sure what the differences, if any, between the two doctrines actually are. I shall address two interpretations of the management–by–results doctrine.

Characteristic of results–oriented management is that the principal early on signals to his agents that the results will be controlled after the fact. Also he indicates what results he expects them to achieve. Furthermore, the superior makes clear that the results will be disclosed to the public and compared to the results produced by other units. To this end, the principal institutes mechanisms for after–the–fact monitoring and evaluation of outcomes and outputs. In management by results, the evaluation mechanism is intended to have a double function. It is intended to give signals to the agents which results the principal is expecting. This is largely to be effectuated by means of the evaluation process, that is, the agents are thought to be influenced by what the questions posed in the evaluation reveals about the value criteria and value standards of their superiors. In addition, the evaluation results are also intended to furnish the principal and the agents with continuous and systematic feedback of reliable information concerning real results. The idea is that the agents themselves voluntarily will correct their activities when they get to know their own position in relation to others. For the evaluation will cover not only one unit but many, which makes comparisons possible. But the reporting of the results shall also be an instrument for the principal to give signals to the agents in which directions they ought to change their efforts. Not only may these signals come as advice and recommendations but also in the form of increased or decreased appropriations and funds.

Obviously, evaluation is a fundamental feature of results–oriented management. Naturally, the results accounted for in the evaluations



are important. Management by results is based on the notion of continuous and systematic feedback to principals and agents of reliable information on policy and program results. Public policy and planning should, above all, be grounded in a firm foundation of real results. By results is meant either outcomes or outputs, or both. Management by results is an outgrowth of a just–give–them–the–facts philosophy. Just present the facts, but leave it to the principal and his agents to apply criteria of merit to the facts and perform the overall evaluation. On the basis of this, principals and staff at all levels in the organization can start to calibrate the goals and objectives for the continued program activities. But it is also important to note that in management by results not only the evaluation findings, but also the evaluation process, is supposed to give critical clues to agents for their future activities.

One similarity between management by results and management by objectives is that the agents are given full freedom to choose means to reach the signalled goals. The principal provides unequivocal result–oriented comprehensive goals, but does not interfere with how the agency or the unit act to accomplish these goals. Management by results is thought to function largely through the force of example. Subordinates should be stimulated by comparing themselves to others, not by being mandated to perform specific tasks through detailed compulsory regulations and directives.

What is meant by results in results–oriented management is not always clear. According to one interpretation, results means substantive results minus resources consumed, that is, costs are taken into consideration. But results also seems to refer to outcomes, outputs, or both without any specification of the costs involved.

According to a second, somewhat different rendering, results–oriented management is regarded as the incipient stage in the development of a management–by–objectives system. Before clear and determinate objectives can be worked out on higher levels, lower–level agencies must account for their actual results. When this is accomplished, higher levels can start to set realistic and feasible goals for future action. Results–oriented management will then become a process that starts bottom up, while management by objectives starts top down, at least in the incipient stage.

The reformation of the Swedish government budgeting process currently under way is dominated by management–by–results rhetoric.

Dissatisfaction with traditional budgeting has been growing for a long time. Critics have argued that budgeting to a preciously small extent is concerned with achieved results; the result–oriented evaluations that are carried out lead lives of their own, isolated from the budgeting process, which mostly concerns appropriations to different purposes. The present reform is an attempt to remedy this unhappy situation by tying output and outcome evaluation stronger to budgeting.

Agencies are placed in a triennial cyclical pattern, so that each of them is subjected to extensive assessment every three years. On these occasions, the assessment should be strongly results–oriented, which presupposes that the agencies themselves continuously evaluate their own programs and disseminate this information to their superiors. In the years when no extensive assessment is performed, the agencies will be subjected to a so–called simplified assessment. Even simplified assessments will contain an evaluative component, since their purpose is to monitor whether the results objectives set in the extensive assessment are met. Furthermore, the government should define comprehensive output and outcome objectives and provide some general guidelines concerning agency activities, but leave the agencies considerable leeway to choose the appropriate ways and means to attain the results objectives. The focus will also be switched from budgeting and other constraints toward results analysis, evaluation, and achieved results (Sandahl 1992; P 1987/88:150, app. 1, FiU30, rskr394, SOU 1990:83).

The attention that management by objectives has received in scholarly journals has seen a dramatic demise. Extensive discussion of the subject could be found in any public administration outlet during the 1970s; articles are rare in the 1980s and 1990s. After the initial euphoria, it has also attracted massive criticisms from the academic community. In government it is impossible to manage by goals or results, it is argued. Politicians cannot determine clear goals because they lack necessary knowledge or hesitate to reveal their operative motivations because they are regarded as inopportune. Furthermore, the official goals are often too hazy to be employed as meaningful criteria of evaluation. A more principal criticism asserts that some activities, such as academic research, simply should not be managed by goals or results. Academic research is an innovative enterprise that cannot be managed by goals for the obvious reason that it should produce something new, which is not known at the moment when goals should be



set. It is also alleged that decisions on means cannot be left to the bureaucrats, because means are sometimes more politically contested than goals. Often, political parties hold widely differing views about means because they are immediate and concrete, while they agree on the goals, which are regarded as cosmetics to garner acceptance and legitimacy for the programs. For instance, it is difficult to see the choice between private or public day care for children as an empirical question of best available means to reach a commonly approved goal: best possible care for the young generation. On the contrary, in cases like this, political parties agree on the goals but widely disagree on the means to reach the goal.

## Note

1. The historical development of retrospective evaluation in various countries is charted in works by Derlien 1990a and others in Ray C. Rist 1990, by several authors like Nioche, Gray and Jenkins, Leeuw, Kordes, and Sandahl in Mayne and Bemelmans–Videc et al. 1992, and by contributors to the volume edited by Gray, Jenkins, and Segsworth 1992.

# 4

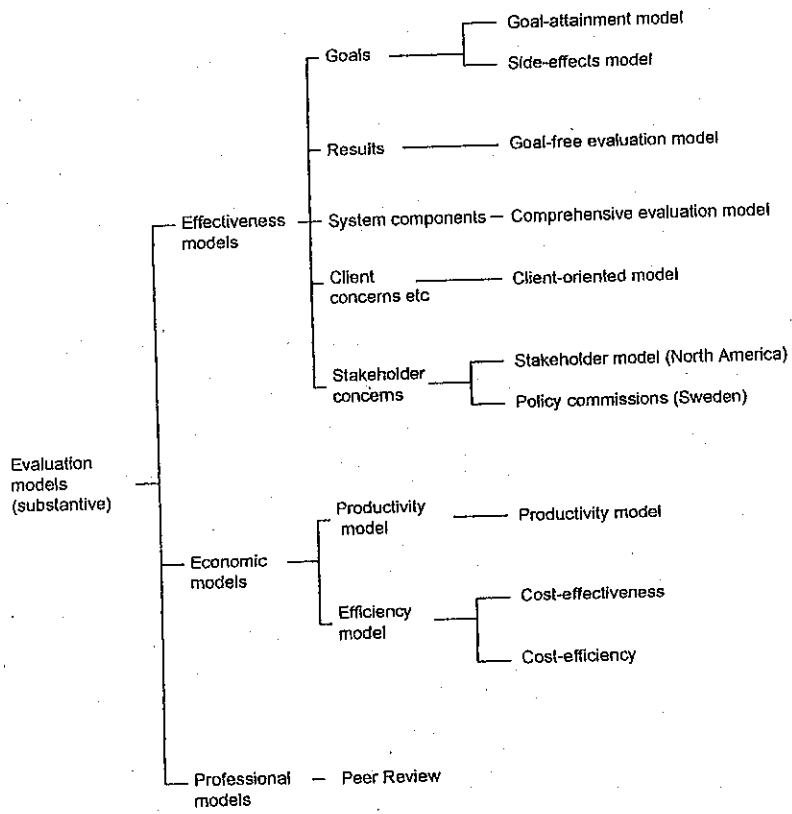
# Models of Evaluation

There is acrimonious disagreement on the methodological foundation and practical orientation of evaluation research. Scores of advocated approaches have been sufficiently formalized to appear in the literature. To create an overview, this multitude of approaches ought to be collapsed into a few basic schools. For this an Occam's razor is needed, but unfortunately no such sharp, universally recognized instrument is available. The most pertinent basis of division is the evaluation's "organizer", that is, its logical point of departure. The organizer is the basic question posed in the evaluation. Often, the chosen criterion of merit is the organizer of an evaluation. But since this is not always the case, I prefer the more abstract concept of organizer.

My approach is based on what Guba and Lincoln (1981:8, 11ff.) have suggested about organizers in their treatise on education evaluation. Since I shall attempt to apply both a political science and a broader social science perspective on evaluation, my scheme will in practice be considerably different from theirs.[1]

A taxonomy of evaluation models, ordered according to organizers, is provided in figure 4.1.

An important line of demarcation distinguishes models focusing on the substantive results of government interventions on the one hand, and models checking for legality, equity, representativeness, and other qualities of the procedures according to which the interventions are supposed to be handled by ministries and agencies on the other. The former are called substantive, the latter procedural models. While of long-standing importance, procedural models will not be dealt with




here; for that reason, they are not entered into figure 4.1. Our study will focus entirely on some important substantive models: effectiveness models, economic models, and professional models. Effectiveness models constitute a fairly large and varied group. In addition to the classic goal–attainment evaluation, effectiveness models include side–effects evaluation, goal–free evaluation, comprehensive evaluation, client–oriented evaluation, and the stakeholder model.

While effectiveness models address only the results of alternative interventions, economic models pay heed to costs as well. The latter property differentiate them from effectiveness models, which are cost blind. The two basic variants among the economic models are the productivity and the efficiency models.

Economic models integrate cost and effectiveness aspects of public interventions. Yet just as evaluators often consider only the results



aspect (effectiveness), administrators sometimes deliberate only the costs. In the 1980s, in fact, there has been an increasing emphasis on the management of resources rather than the management of policy. However, pure cost models are not included here.

Before proceeding, a caveat must be inserted concerning "effectiveness" and "efficiency." In efficiency measurement costs are included, while effectiveness assessment concentrates on results only without taking costs into consideration. This was clear to Herbert A. Simon already in 1945 when he wrote his chapter "The Criterion of Efficiency" in his renowned book *Administrative Behavior* (1976:180f.). "Until practically the end of the nineteenth century, the terms 'efficiency' and 'effectiveness' were considered almost as synonymous," Simon asserted. But then, efficiency acquired a second meaning: the ratio between input and output. To sustain his assertion, Simon quoted *The Encyclopedia of the Social Sciences*:

> Efficiency in the sense of ratio between input and output, effort and control, expenditure and income, cost and the resulting pleasure, is a relatively recent term. In this specific sense it became current in engineering only during the latter half of the nineteenth century and in business and in economics only since the beginning of the twentieth.

Consequently, intervention effectiveness in English has nothing to do with costs.

Professional models, the third basic category in figure 4.1, focus on the subject matter only indirectly, in that immediate stress is put on who should perform the evaluation. The most celebrated professional model is the peer review approach, in which for instance professors evaluate professors, engineers, and surgeons.

My exposition will start with effectiveness models, proceed through economic, and end with professional models. As already mentioned, the procedural models will not be treated in the present context.

## Goal-Attainment Evaluation

The classical way of approaching the evaluation problem is goal-attainment evaluation. The two basic ingredients of goal-attainment evaluation are called goal–achievement measurement and impact assessment. The key question in *goal-achievement measurement* is: Are the results in accord with program goals? And the *impact assessment*

issue can be formulated: Are the results produced by the program? Another name for goal–achievement measurement would be *results monitoring*.

Goal–attainment evaluation is a paragon of simplicity and lucidity. After identifying the goals of the program, teasing out their actual meaning and rank order, and turning them into measurable objectives, the second step involves determining to what extent these premeditated goals have been realized in practice. The third step in goal–attainment evaluation implies ascertaining the degree to which the program has promoted or dampened goal realization.

Goal–attainment evaluation is an effectiveness model because it asks questions about the substantive content, output, and outcomes of the program, not about program procedures like equity of treatment, due process, and the like. It differs from economic and institutional models in that it raises substantive issues only, but pays heed to neither program costs, nor the organization of the evaluation.

In adopting the yardsticks of others as criteria of merit, the goal–attainment model applies a descriptive theory of valuing. Actually, it supports a particular descriptive theory of valuing, since it takes premeditated program goals as criteria of merit and organizer for the evaluation. Ernest House writes that

> [it] takes the goals of the program as stated and then collects evidence as to whether it has achieved those goals. The goals serve as the exclusive source of standards and criteria. The evaluator assesses what the program developers say they intend achieving. The discrepancy between the stated goals and outcomes is the measure of program success. (1980:26)

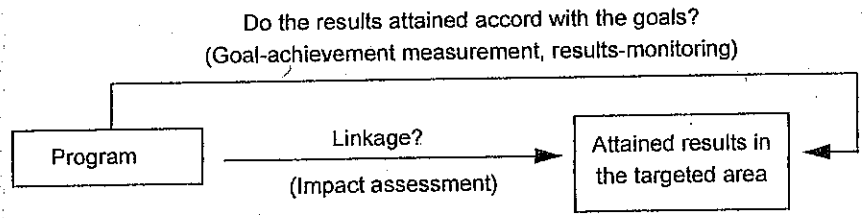
This does not imply, in practical terms, that evaluators should first of all search out the goals. The crucial point is that goals are the *logical* point of departure. After all, the major task of evaluation, advocates of the goal-attainment model claim, is to determine if the premeditated program goals have in fact been achieved and then try find out to what extent the program has contributed to goal achievement.

The simple anatomy of goal-attainment evaluation is outlined in figure 4.2.

The first part of the goal-attainment model, goal–achievement measurement, engenders two distinct activities to be kept apart: (1) the clarification of program goals (the goal function); and (2) the measurement of actual completion of premeditated program goals (the goal accomplishment function). The second part of the model implies finding out to what extent the program has contributed to goal achievement (the causal function) (Lane 1987).



**FIGURE 4.2**
**Goal-Attainment Evaluation**




Let me provide a straightforward example. The goal of the 1978 Swedish Energy Savings Plan for Existing Buildings was that "annual net energy consumption" in the 1978 building stock would in 1988 be approximately 35 TWh lower than in the benchmark year of 1978. This meant a reduction of approximately 30 percent. To achieve this goal, state funds were channelled into home insulation, retrofitting of industrial and commercial buildings, and energy auditing and counselling. A goal–attainment evaluation of the Energy Savings Plan would entail investigating whether annual net energy consumption—however measured—in the 1978 building stock actually was 35 TWh lower at the end of 1988 and whether this actual outcome might be attributed to the adopted reform.

The example makes it clear that the goal–attainment model also presupposes some idea about the time span that may pass before the goals are supposed to be achieved. In other words, goal–attainment evaluators must know when it is appropriate to pass a judgment on goal completion. In my example, a full summative evaluation can only be carried out in 1989 or later. However, goal–achievement evaluations can also be performed as intermediate reviews during this period to check whether things are moving in the right direction.

The goal–attainment model may be applied to government interventions at all political and administrative levels; for example, programs adopted by:

1. municipal commissions;
2. municipal assemblies;
3. county assemblies;
4. national agencies;
5. national parliaments;
6. the Nordic Council;
7. the European Community; or
8. the United Nations.

A program might be a broad regulatory mandate espoused by a United Nations body or by a national parliament, or tiny planks in small substantive programs introduced by a municipal decision–making body. Embraced by most OECD countries and truly global in magnitude, the 1987 Montreal Protocol regulating emissions of substances thought to destroy the global ozone layer is a program, as is the peat heating scheme enacted by the city of Uppsala in Sweden.

All types of government programs can be evaluated for goal attainment. Programs may concern organizational change in the public sector or substantive issue areas. A program can be many different things: a planned public social service, an actual social service, a stream of commodities, a law, a norm derived from a law, a planned information effort, and the like. The intervention concept also covers conglomerates of individual programs, for instance Swedish environmental policies, which include numerous important laws, subsidy schemes and grants for research. Neither the kind of program nor the political or administrative level at which it is instituted is of importance to the subsequent reasoning on goal–attainment evaluation.

While "goal-attainment evaluation" is a common expression in the literature, the term "goal–achievement evaluation" is frequently used as well. Other denominators include "the rational model," "the objectives-oriented approach," and "the behavioral objectives approach."[2]

## The Strength of the Goal–Attainment Model

In earlier literature, the goal–attainment model reigned supreme. Program evaluation was even identified with goal–attainment appraisal, a curious fact that was discussed in chapter 1 on the evaluation concept. Since the 1970s, however, goal–attainment evaluation has been under constant attack. Presently, almost no one is able to, or has the stamina to stand up in its defense.



There are, however, three important reasons in favor of goal–attainment assessment. I shall label them the democratic argument, the research argument, and the simplicity argument.

The *democratic argument* is a compelling one, grounded in the notion of the *primacy of the parliamentary chain of control* and, consequently, in a democratic perspective.

Government program goals differ from goals in other walks of social life due to their institutionalization through a formal decision–making machinery. They are publicly and officially adopted in political assemblies by the representatives of the people. Decision–making procedures of political bodies are circumscribed by formal rule systems to an extent that has no counterpart in, for example, corporations, voluntary associations, or families. Political leaders are supposed to honor the rules of the constitution, and the rules of procedure in parliament and government. Once a decision comes out of this system, it has a status that cannot be compared to decisions in other social bodies. The fact that program goals are the product of the formal decision–making machinery of the state makes them all the more significant. It is a merit of the goal–attainment model that it recognizes this fact, which is often overlooked by students of general organizations.

The goals of the democratic state also occupy a special position because they have emerged under responsibility. In principle, political officials must contemplate not only their own wishes, but also available resources. Partial interests, like the various pressure groups, can pursue demands and set goals without considering the general public interest or the financial situation.

In the extension of this conviction of the special status of the parliamentary chain of command there is a deeply rooted ideal of public administrators as the obedient tools of the elected representatives. Civil servants must attend to their political principals regardless of personal idiosyncracies, feelings about the worth of the programs, and partisan allegiances. In serving their masters, they should subscribe to Tacitus' famous *maxim sine ira et studio*, "with neither anger nor partiality."

In sum, I am sympathetic to the descriptive theory of valuing underlying the goal–attainment model.

Two of the criticisms periodically raised against the goal–attainment model are insensitive to the democratic aspect. One of them brands the model as dangerously *elitist*. It is geared to the needs of reform makers and other peak groups in the decision–making system,

since it is these power wielders who want to know whether program goals have been accomplished.

The elitist criticism disregards that goal accomplishment is of the utmost importance from a citizen's perspective. It must be of interest to citizens to be informed about whether the adopted policies really deliver what they promise. Assume that the representatives of the citizens, the elected politicians, have promised in an election to introduce a pension reform that will give everyone an economically secure retirement. It eventually becomes obvious that these goals are not being fulfilled. It must be valuable for the electorate to get to know about this when they are to judge the performance of their representatives, and past policies in the next election.

The second objection that overlooks the accountability and citizen perspective intimates that initial program goals might become *obsolete* as evaluative yardsticks. In adopting initial goals as points of departure, the evaluator runs the risk of using an antiquated baseline when her paper is presented. The result cannot be applied in the next round of decision making.

The compelling force of this objection hinges upon which goals that change and to whom the evaluation is addressed. If goal displacement occurs among politicians, evaluation against official, allegedly obsolete goals will be of no interest to them. But it will be of great concern to the citizenry in its role as principal, because to them the official goals still hold true, and they may want to hold the politicians accountable for discarding them. If goal displacement subtly transpires in the national agency or even further down in the implementation process and the program is modified radically due to this, evaluation against antiquated official goals will be of no substantive concern to these groups, but will be of the utmost relevance to citizens and responsible politicians, since agencies and low-level implementors are their proxies and executives, whom they may want to hold accountable.

In sum, the goal–attainment model scores important points with respect to its inherent descriptive theory of valuing, which tilts toward the parliamentary chain of control, and representative democracy, particularly toward the frog's-eye perspective of the citizenry. The citizen perspective has been repeatedly stressed in this context, because it is rarely alluded to in international discourse on public program evaluation.

The second reason for goal–attainment assessment, *the research argument*, is based on the model's potential for bringing social research to bear on the evaluation enterprise. All evaluation requires dimensions of merit through which the program can be graded. Evaluation is a normative enterprise. The goal–attainment model seems to offer an objective solution to the value criterion problem in evaluation. Since program goals are explicitly stated in the original legislation or in the preparatory work, they can be established empirically, through an act of interpretation. By judging program results from program goals, the cautious evaluator can avoid taking a personal, subjective stand on the merits and demerits of the programs to be evaluated. Goal–attainment evaluation embraces a descriptive theory or valuing; it is what Scriven (1991:30) has called a secondary evaluation—evaluation from the point of view of others. Since the criterion issue can be settled in an objective fashion the whole evaluation can be conducted in an objective fashion as well. The goal–attainment model accords a large role to the objective evaluator. There exists a group in society who are trained to conduct objective research: the academic social scientists. Evaluation is best performed if the task is assigned to independent social scientists.

A third, admittedly less important reason for the goal–attainment model is its attractive *simplicity*. Involving only two major questions, it is very easy to understand and apply.

On the other hand, the goal–attainment model also suffers from persistent flaws and weaknesses, to which I shall now direct my attention.



## The Shortcomings of the Goal–Attainment Model

The goal–attainment model disregards *costs*.[3] Goal accomplishment may have incurred substantial sacrifices in terms of money, time, and human efforts; these are completely ignored in goal–attainment assessment.[4] The lack of cost–consciousness alone reveals that the goal–attainment model cannot aspire to be the sole valid model of public policy evaluation.

Even more damaging is the case against goal–achievement evaluation in its own purported area of competence. One such momentous argument suggests that program goals are deficient as criteria of merit because of their *haziness*.

While a truism, the supposition of the pervasive haziness of policy

and program goals is repeatedly stated in policy analysis literature. So accepted is the general idea that only rarely efforts are made to indicate in what ways public sector goals are muddy. There are good reasons, however, to distinguish between two major types of goal obscurity: goal indeterminateness and goal catalogues.

First, programs may be based on *indeterminate goals*. There is "terminological inexactitude," to use Winston Churchill's famous expression. Terminological inexactitude, in turn, is of two kinds. Particular goals may be *ambiguous* and carry two or more simultaneous meanings. Yet ambiguity is probably quite exceptional in political and bureaucratic language, and barely bothers evaluators. More confusion is caused by *vagueness*. A goal is vague if it does not delineate clearly cases where it is or is not applicable. The outer border delimiting the extension of a vague word is so fuzzy that within a certain range it is impossible to know what is included in the extension and what is not. Rampant in political rhetoric, vagueness is one favorite expedient to settle political conflicts without really resolving them.

The second major obscurity is produced by *goal catalogues*. In connection with most large social reforms, impressive directories of diverse goals are regularly presented. While a single goal may be hailed as the major one, it is often also said that this one must be balanced against all the others, maybe including potentially conflicting ones. Often, the particular items in the goal catalogues are vague. In addition, the necessary trade–offs between the several goals are not indicated, which makes it impossible to elicit from such lists of goals one distinct, transparent, expected outcome. Thus, program goals do not offer any safe guidance for the direction of continued empirical research work. They are not lucid enough to be usable as value criteria against which to measure successes, shortcomings, and failures.

The government regulatory regime concerning Swedish forestry management contains a catalogue of individually vague and collectively contradictory goals. The first paragraph of the 1979 Forest Management Act states that forests must be managed in such a way as to secure their sustained capacity to produce timber. However, the law also declares that nature protection and other societal interests must be paid heed to. The forestry business must operate with due concern for the functions of the forests as habitats for plants and animals, their role in water balance and local climate, for outdoor life and recreation. Attention must be paid to valuable cultural environments and landscapes.



Exactly how sustained high production of timber should be balanced against the other goals is not stipulated. This is probably wise, since it would be difficult to specify in central decisions how the weighing should be made in every local case. However, this wisdom creates difficulties for the goal–attainment model, since the program does not stipulate a clear expected outcome.

If the evaluator sticks to the notion of objectivity, which is a fundamental tenet in goal–attainment evaluation, she will not be able to tease out from the overall goal formulations one indisputably clear global outcome objective. To arrive at such an objective, she probably has to clarify the program goals and make priorities among them, which entails leaving the sphere of objective social inquiry and entering the area of subjective speculation.

The goal–haziness argument reveals an important misfit between the requirements of the goal–attainment model and the way public policies and programs are usually composed. If elected officials and program planners have not specified individual goals into measurable objectives, and if they have not balanced the various stated goals into one, global outcome or output measure, the goal–attainment evaluator cannot sum the results into a final evaluative judgment in any completely objective fashion. She can do so only after she has clarified the goals and prioritized among in a fashion that will cast doubts on the objectivity of the whole enterprise.

On the other hand, the goal–haziness argument, while pertinent, is not applicable across the board. Occasionally, political bodies find it appropriate to set clear, even quantified objectives; an obvious example is the above–mentioned 1978 Energy Savings Plan. In these cases, the goal–haziness argument has no bearing on the full applicability of the goal–attainment model.

A third objection, about *unintended effects*, is in my view the crucial one. Attempts at deliberate political control invariably lead to consequences that were not originally foreseen in the original decision situation. "It is difficult to forecast, particularly about the future," as Niels Bohr jokingly put it. Were the evaluators to confine themselves exclusively to researching the achievement of premeditated program goals, any serendipitous results or unanticipated side–effects would not be included in the main evaluation process. The evaluation would provide a tunnel vision of events, and produce a biased, if not fundamentally wrong, picture of what the program has attained. In all likeli-

hood, a program generating some interesting spin–off effects must be better than a program producing several undesirable spillovers (Foss Hansen 1989:204).

In addition, the prevalence of intentional action resulting in unintended side–consequences is one of the strongest reasons—besides perverse effects—for performing evaluations in the first place (Meyers 1981:18ff.). Designing them in a way that does not allow for the possibility of discovering such side–effects must then be a serious mistake.

Suppose somebody made a goal–attainment evaluation of King Ferdinand and Queen Isabella's program of sending Christopher Columbus westward to find the sea route to India. Our evaluator concludes that the program is a failure because India was not reached and the program goal was not achieved. There is no reason to consider the side–effects, the discovery of America, the ensuing cruel and deceitful extermination of advanced ancient civilizations, and so on and so forth, since the goal–attainment model pays no heed to phenomena outside the target area.

A fourth seemingly compelling objection suggests that the goal–attainment model disregards the role of *hidden agendas* in public policymaking.

It has been suggested that policy analysts using premeditated, officially pronounced, substantive goals as measuring rods for evaluative purposes and organizers for data collection disregard something important in the political game situation. Manifest substantive goals resemble the tip of an iceberg. Far from revealing the real operative motives, they constitute only what decision-makers want to hold up to the public. Officially stated goals may have a symbolic character, not intended to be achieved, while real, hidden motives point to other directions. The hidden substantive political agenda behind a privatization drive of government–owned companies may not be to increase efficiency in the best interests of the consumers but to strengthen the power network supporting conservative and other nonsocialist parties. Hidden agendas include strategic motives as well. The point may be to strengthen party cohesion, to keep or attract voters in the next general election, or to prepare the ground for a coalition government. Once these hidden strategic goals are achieved, politicians lose interest in implementing the substantive provisions of the program.



To political scientists, this is a staple analysis of the motivating forces prompting public reforms. Of course, strategic considerations play an important role in public policymaking. But what is the real import of the hidden–agenda stricture against the goal–attainment model? Should the results also be evaluated against such strategic yardsticks as winning the upcoming election, strengthening internal party cohesion, or maintaining government coalitions? In that case, we have to admit, the goal–attainment model would be insufficient since it uses substantive and only substantive goals as yardsticks.

On the other hand, goal-attainment evaluators may rightfully argue that hidden agendas may very well be considered in their model. They might be used as factors *explaining* why substantive results did or did not occur. If, for instance, substantive program goals were not met, there is nothing in the logic of the goal-attainment model impeding evaluators from offering hidden strategic agendas as reasons why this occurred. Providing such explanations must be considered a major activity in evaluation research, and it seems to be reconcilable with a slightly widened goal–attainment model.

On the other hand, a strictly applied goal–attainment model does not emphasize nonprogrammatic explanations of the results. There is no separate box in the goal–attainment model for strategic considerations as explanations of outcomes. In fact, there are no boxes except for the one signifying the program. Naturally, the goal–attainment model must consider other explanatory factors besides the program but they are lumped together into one, large comprehensive factor, which is contrasted with the program factor. This comprehensive factor is treated like a "black box," never to be opened, but this is an obvious drawback, since if the program was partly a failure it would be very interesting to know exactly why.

In conclusion, the goal–attainment model has problems in considering hidden agendas underlying program decisions, but these problems are not insurmountable.

A fifth counter–argument maintains that the goal–attainment model *does not regard implementation as a problem.* Implementation processes are treated as black boxes. This is also a pertinent stricture, since if strictly applied according to figure 4.2, the model does precisely that. The model concentrates on the consequences of the program, not of the implementation processes. The organizations and networks of institutions and actors which are supposed to transform

decisions into effects are uninteresting to goal–attainment evaluation.

A last objection maintains that the goal–attainment model proceeds from an all too *conventional view of the politics–administration relationship* according to which the administration not only in theory but also in practice promptly and faithfully executes the decisions of politicians. Implementation processes are actually mechanical, straight line, and controlled from the top. The goal–attainment model is accused of regarding public administration as a target–seeking robot fulfilling its mission with painstaking technical precision.

But this disparagement entirely misses the point. The goal–attainment model may well be associated with the normative conviction that the civil service ought to be the faithful tool of the elected representatives of the people. It does not, however, take for granted that civil servants always loyally execute the decisions of the politicians. There is nothing in the logic of the goal–attainment model compelling us to believe that the program has produced the intended results. On the contrary, eventual program impact in the target area is considered one of the two problems to be investigated in a goal–attainment evaluation, the second one being whether the achieved result accord with the original goals. The arrow from "program" to "attained result" in Figure 4.2 shows, not how the linkage actually is, but what the evaluator attempts to test in a goal–attainment impact assessment. She may very well come up with the conclusion that there is no causal relationship between program and result. The objection that the goal-attainment model treats programs as target–seeking robots has no bearing whatsoever on the logic of the model.

In sum, taking official subject–matter goals seriously is the major strength of the goal–attainment model. The reason why this is a strength is grounded in the theory of representative democracy. From a governance perspective, citizens as well as elected politicians need goal evaluation to check whether their executives actually carry out what they are obliged to do.

However, the goal–attainment model is also open to several valid objections. It overlooks procedural goals like equity of treatment and legality. It disregards costs. It has some problems with hazy goals, pervasive as we all know, in public policymaking. The most compelling rebuttal, however, emanates from the fact that the goal–achievement model is blind to side effects.

At this point, I would like to present a model that expressly considers the weighty side–effects argument, while retaining the fundamental goal–orientation of the goal–attainment model: side–effects evaluation.



## Side–Effects Evaluation

An inherent difficulty with goal–attainment assessment using prespecified goals as the organizer is how to account for unintended and unrecognized effects. Political actions may culminate in unintended, unforeseen by–products. If an important set of side effects is unintended, how can they be considered and judged if the evaluator sticks to premeditated, substantive goals as the basic organizing principle? To solve this problem, the goal– attainment model must be extended to cover side effects. I shall call this approach side–effects evaluation.

Characteristic of the side–effects approach is that goals are retained as the fundamental organizer but supplemented by side effects. For the expression "side effect"—a side effect in relation to what?—presupposes knowledge of what the main intended effect is expected to be. Side effects must be defined in relation to the intended main effects. From the policy instigators' point of view, a side effect may be defined as an impact outside the program target area. "Main effects" can be defined as the central substantive impacts that the policy instigators *by intention* wanted to achieve. Consequently, main effects are associated with the substantive objectives of the policy–makers and with what they believe themselves capable of achieving. Furthermore, main effects are by definition anticipated as well as positively valued by the policy instigators.[5]

Thus, I contend that the model is *based* on goals in that program goals are retained as the focal organizer. But in addition, traits outside the target area preordained by program goals are also considered. The underlying idea is that public interventions may produce other things than intended results. They may lead to great discoveries embodied in names like Columbus and America. But they may also create more problems than they solve, turning solutions into problems. Solutions to puzzles become puzzles requiring solutions.

The side–effects issue is still grossly overlooked in evaluation research. I agree with the judgment of Sieber (1981:44f.): "Concern with unanticipated consequences of any sort seems to be a grudging

exception [in evaluation research]. Textbooks on evaluation rarely mention the subject, and when they do it is almost entirely in the form of sheer admonition (along with a footnote reference to Merton, perhaps). Concrete guidelines for conceptualizing, detecting, measuring, or assessing a net balance of good and bad effects are not offered."

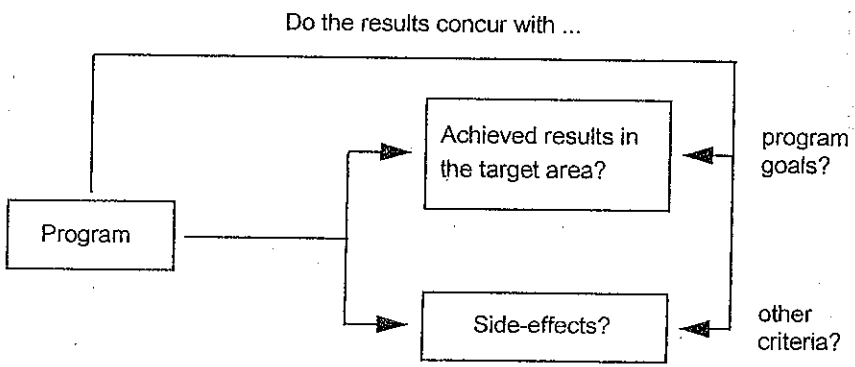
The basic skeleton of the side-effects model is sketched in figure 4.3.

It has been known since Machiavelli that political actions can be counterproductive. The liberator easily turns into an oppressor—the strength needed to break the bonds will be used to again put people into bonds, which in turn must be broken. This is what is known in public policy as *perverse effects*.

Perverse effects run exactly counter to the very intentions of the program instigators. To use a different but pertinent terminology, they are cases of counterfinality. Since these impacts occur in the target area of the public intervention they are not side effects. Neither are they main effects, since they are not coveted by the policy instigators. The side-effects/main-effects terminology cannot catch perverse effects. Perverse effects of purposive action ought to be treated as a separate category.

Perverse, or reverse, effects are also different from *null effects*. According to Sam D. Sieber, who to the best of my knowledge origi-

**FIGURE 4.3**
**Side–Effects Evaluation**






nated the distinction between perverse effects and null effects, which I am applying here (1981:10, 47, x), the expression null effects means that programs adopted because they are earnestly and fully expected to have certain effects wholly fail to spawn any such effects. The programs produce no impacts at all on their targeted areas. In the perverse effects case, consequences are produced but entirely contrary to the ones intended.[6]

A nearly universal fascination with the topic of perverse unintended effects of intentional action is manifested by an array of technical and historical terms—the Cunning of Reason, the irony of history, Pyrrhic victory, two–edged sword, counterproductive measures, backlash, and boomerang effect.

An alleged classic example (Sieber 1981:9, 166) of a self–defeating public policy is the Maginot Line, which lulled the French into a false sense of security and the consequent immobilization of military resources which contributed to their defeat by the German army in 1940.

Since by definitional fiat perverse effects occur and null effects do not occur in the targeted areas, the goal–attainment model with all its attention directed at what happens in these particular fields has no problem with handling them. But this also means that the model cannot discover and ascertain side effects because they fall outside of the targeted sectors.

Side effects can be *anticipated* and considered in calculations preceding decisions to adopt policies. Side effects are by far not always detrimental either. They may be detrimental to policy instigators, but, of course, beneficial as well. In spite of the fact that the term *side effect* has a slightly derogatory connotation in English, the evaluator must be open to the fact that public interventions may produce welcome as well as unwelcome side effects.[7]

In the aftermath of the 1973 oil crunch, governments all over the world instituted energy conservation programs. In Sweden, a government grant program was adopted for energy conservation in dwellings. Later it was converted into the above-mentioned Energy Savings Plan for Existing Buildings (from Vedung 1982a:85). State subsidies were offered to people willing to retrofit their properties. The intended outcome was to achieve more efficient energy use in existing buildings. Yet it was anticipated that the grant program might have an unfavorable impact on the distribution of wealth in the country, since home owners, owners of multifamily dwellings, and other prospective re-

cipients of the program are generally wealthier than people at large. While not primarily coveted, policy-makers were prepared to create and accept this side effect in order to reach the overarching desired goal of energy efficiency.

On the other hand, policy sponsors also realized that the subsidy scheme might spawn a number of felicitous spillovers. It would boost economic activity in general and create numerous new jobs. Retrofitting and insulation of walls and attics would diminish draft, which would make indoor climate more comfortable. Insulation and the installation of three–pane windows would reduce noise.

A study geared toward investigating energy conservation as well as wealth distribution, employment, and comfort aspects would be a side–effects evaluation.

So far, I have considered anticipated side effects only. However, some side effects are no doubt *unanticipated* as well. Some policy analysts would probably even argue that most side–effects are unanticipated at the time the decision is taken. Like their foreseen counterparts, they might be felicitous or deleterious. Increased radon radiation in dwellings, and the increased incidence of allergies, both probable consequences of better insulation causing less draft, may be adduced as examples of deleterious, unanticipated side effects.

Beneficial inadvertent consequences are rare specimens, however, because reformers trying to sell a new policy are likely to list and exhaust all the positive impacts likely to follow. "The phenomenon of overselling a program at the time of policy formulation and legitimation is well-known and, in fact, characterizes much of policy–making in the United States," argue Ripley and Franklin in a recent book on implementation (1986:234f.). However, there are occasions when grim policies have unanticipated, happy spinoffs.

The demilitarized zone around the 38th parallel formed in 1953 across the Korean peninsula is an interesting story of how military conflict resolution unintendedly produced a last refuge for endangered species like, for instance, the rare Manchurian crane (*grus japonensis*), the ringneck pheasant, the native Korean bear, and the wildcat. The 4000 meter wide DMZ is one of the most militarized areas in the world as it is mined, has obstacles to foot or vehicle movement, and is constantly patrolled. It is also one of the least inhabited. It is the militarism of this area that inadvertently has created a safe haven for all wildlife as the habitats are saved from encroachments from intensive agriculture, industry, the building of roads, and the construction of cities, are protected from the industrial pollution that has threatened wildlife in more populated areas, and the animals very seldom find themselves as someone's supper.



The discovery that social phenomena may be "the results of human action but not of human design" is usually accredited to Bernard Mandeville and particularly to the moral philosophers of the Scottish Enlightenment such as Adam Smith and Adam Ferguson.[8] In particular, they predicated their case for the free, unfettered market on the idea of happy, inadvertent side effects. Mandeville's famous formulation "Private Vices, Public Benefits" refers to the market mechanism's capability of transferring individual evil egoism into collective welfare. The beneficial social results ensue unplanned, as side effects of the egoistic, self–serving behavior of individuals.

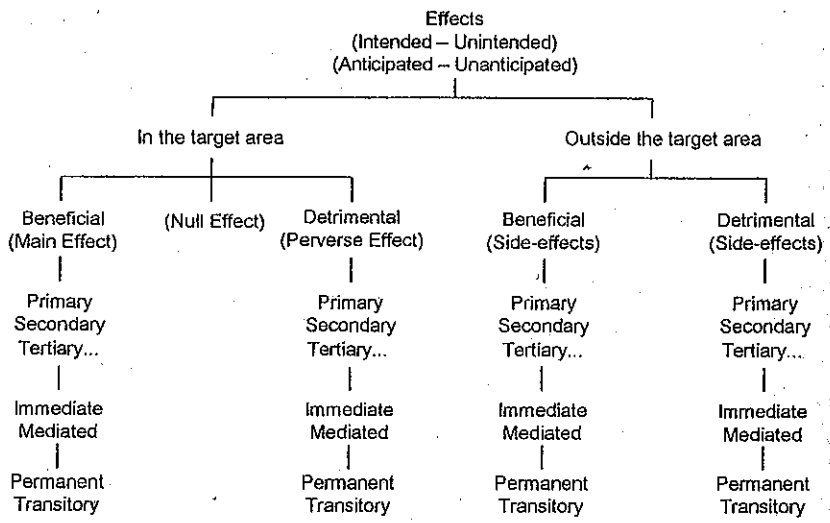
The central concept fashioned by Adam Smith in his *Wealth of Nations* (1776, 1937:423) was the "invisible hand." In an illustrious passage, Smith argued that each individual pursuing his own selfish interest would most effectively, but unintendedly, promote the common weal:

> Every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. . . . .By directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. . . . .By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.

My argument on main effects, perverse effects, null effects, and side effects is summarized in the effects tree in figure 4.4, showing which aspects of effects that might be studied in evaluation research.

A more extended real-life example may clarify my line of reasoning. Aside from reduced energy use—the main substantive effect intended—the Swedish 1974–1980 government grant program to energy conservation in industrial buildings and processes might have produced the following main, perverse, and side effects:

**FIGURE 4.4**
**Main Effects, Null Effects, Perverse Effects, and Side Effects**




| | |
|---|---|
| 1. Main effect (substantive) | Reduced energy use |
| 2. Null effect | Grants are allocated to actors who would have taken the measures anyway |
| 3. Perverse effect (substantive) | Increased long–range energy use |
| 4. Beneficial side effects | Improved work environment<br>Water conservation<br>Reduced operation and maintenance costs<br>By filling out applications for subsidies, applicants came up with other savings and efficiency measures, which have been effectuated. |
| 5. Detrimental side effects | Destruction of capital assets because process equipment is changed earlier than otherwise would have been the case<br>Rising prices on energy conservation equipment<br>Inconveniences to management and workers in conjunction with rebuilding and installation |
| 6. Other side effects | More power to the National Board of Industry |



Undoubtedly, the grants program for energy conservation in industry produced energy savings. The main intended effect was achieved. However, on some occasions the program might have been counterproductive, actually enhancing energy inefficiencies instead of diminishing them. Additional energy-saving equipment might have conserved obsolete technology and counteracted modernization, which could mean that in the long run, more energy instead of less would be used. This would amount to a long–range perverse effect.

A noticeable thing about the grants program is that it produced both beneficial and detrimental side effects. The various beneficial side effects, like the first detrimental one, are quite self–explanatory. The second detrimental side effect could be ascertained only a long time after the initial investment.

There is a great debate among social theorists concerning counterfinality, by–products, and unintended consequences. Karl Popper, among others, in his *Conjectures and Refutations* (1974:124) has forcefully claimed that perverse effects and spillovers constitute the central problem for social scientists:

> It is one of the striking things about social life that *nothing ever comes off exactly as intended*. Things always turn out a little bit differently. We hardly ever produce in social life precisely the effect that we wish to produce, and we usually get things that we do not want into the bargain. Of course, we act with certain aims in mind; but apart from these aims (which we may or may not really achieve) there are

always certain unwanted consequences of our actions; and usually these unwanted consequences cannot be eliminated. To explain why they cannot be eliminated is the major task of social theory.[9]

The importance of noticing perverse effects should be obvious to every evaluator. If the program produces consequences contrary to its main purpose, there must be something wrong with it. But why is it so vital to pay attention to side effects? Because by–products, whether detrimental or beneficial, are crucial factors in every inclusive judgment of the operation of an intervention. Should it turn out that side effects, which have been known, discussed, and positively valued in advance, have not materialized in spite of the fact that the program has been on the books for the intended period of time, this ought to have consequences for any appraisal of the intervention. If the employment effects of economic subsidies to building insulation are much weaker than was calculated with in advance, there will be even less reason than before to keep the program in its present form, even though expectations about energy savings have come true.

Programs may produce spillovers, which in turn constitute or create fresh problems that must be subjected to novel government programs. Solutions to puzzles become puzzles requiring solutions, which in turn become puzzles requiring new solutions, and so on. Aaron Wildavsky writes: "More and more public policy is about coping with consequences of past politics" (1979:4f.,69ff.). Information about side effects is crucial to any comprehensive assessment of government programs.

Unforeseen side effects are of particular interest to evaluators. The actual consequences of a public policy always turn out somewhat differently from what the agent expected before he took action. This is due to the fact that consequences are produced by the application of the program, not by the assumptions originally guiding the decision to adopt the program. Influencing the adoption of the program is entirely different from controlling its outcomes.

By now, we are far removed from the original goals and intentions underlying the intervention. The emphasis would be on global results, planned and unplanned, intended and unintended. The point is, however, that the evaluator still makes a distinction between anticipated and unanticipated results, which presupposes premeditated goals as an organizer. In this limited sense also, the side–effects model is goal based.



Possible sources of suggestions about side effects would be social science theory and the legislative history of the reform. Of particular importance are those apprehensions held by the opponents of the program proposal at the time of its adoption. Other sources would be any public controversy erupting maybe a few years after the adoption of the program, what would come up in government and parliament in conjunction with reassessments of the program, or what the pertinent officials believe about the program when it has been going on for some period of time.

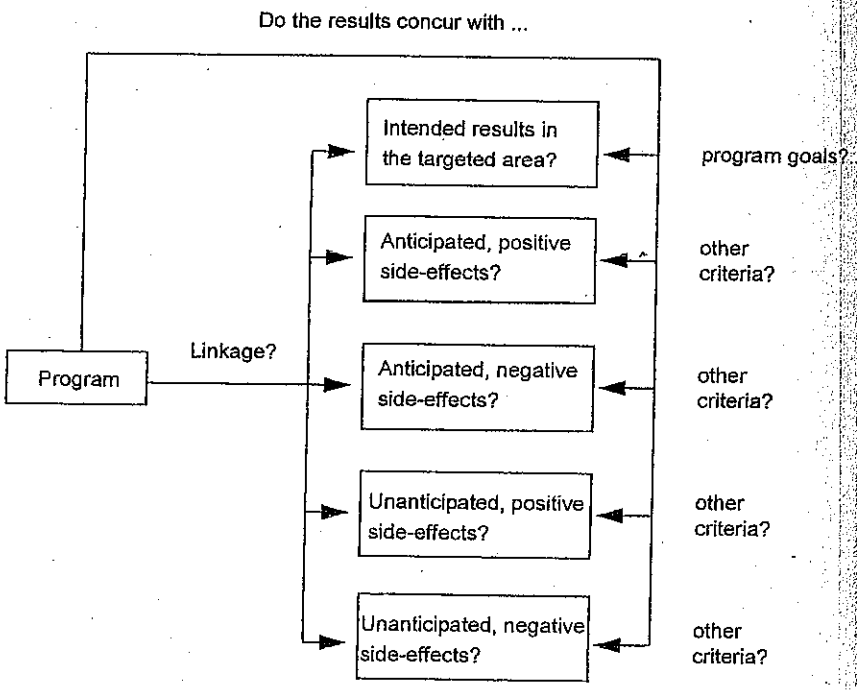
If the totality of effects (positive and negative anticipated consequences, as well as positive and negative unanticipated consequences) of a government intervention were to be investigated, the structure of the evaluation on the outcome side might be as shown in figure 4.5.

Like goal–attainment evaluation, side–effects evaluation is predicated on the normative conviction of the primacy of the parliamentary chain of command. Given this, I strongly prefer side-effects to goal–attainment evaluation. Indeed, the major rationale for doing policy evaluation in the first place is that state actions to some extent are unpredictable and regularly result in side effects not originally foreseen. It is an important duty of evaluation to map and assess the worth of these side effects.

A major challenge to side–effects evaluation is what criteria to apply in judging merit. Ideally, the evaluator would like to trade off the value of the intended main effect of the intervention against the values of the beneficial and deleterious side effects. Negative side effects could then be a grudgingly accepted cost to reach the overall main outcome. Strong positive side effects could also enhance the acceptability of programs with weak goal completion. The performance of this operation requires value criteria for the main effects, for each type of side effect, and for the trade–off between the two.

The descriptive theory of valuing suggested by goal–attainment evaluation—that given policy and program goals should be used as criteria of merit—is clearly insufficient for the following reason. If some effects are not foreseen, the criteria and standards for judging merits and demerits of these effects are not prespecified either. Therefore, goals and other prespecified criteria of merit are insufficient as far as judging each individual unanticipated side effect is concerned. Since the criteria for judging the value of side effects are not stated in advance, the criteria for assessing the trade–offs between the values of

**FIGURE 4.5**
**Side–Effects Evaluation with Specified Side Effects**




the side effects and the main effect are not prespecified either. This will hamper the calculation of the global, aggregated worth of the intervention. To compute the global value, the evaluator must know the value of each side effect individually and add these values to or subtract them from the worth of the main effect.

A feasible solution to this problem involves a different, more creative, descriptive approach to valuing. Aside from mapping the main effect and comparing it with the prespecified goals, she may also chart the side effects but leave it to the commissioners and other users of the evaluation to ascertain their value and carry out the overall, global assessment of the program so to say *ex post facto*. Since the model is predicated on the values of others, it still embraces a descriptive theory of valuing. But the concentration on side effects forces the users of evaluation to consider other values than initially incorporated into the intervention. The evaluator informs the users about a range of effects



which makes it necessary for them to produce and use criteria of merit, which were initially overlooked or not articulated.

## Goal-Free Evaluation

From the difficulties involved with using premeditated intervention goals as data organizers and criteria of merit, at least one bold evaluation theorist has drawn an even more radical conclusion than those hitherto accounted for. I am referring to the goal-free evaluation model, initially designed by Michael Scriven (1973, 1974, 1980, 1991).

At a first glance, Scriven's contribution seems impudent, almost frivolous. His insistence that evaluations should be goal free must be a *contradictio in adjecto*. How could one possibly *evaluate* if one is prohibited from putting up goals to evaluate against? After all, the purpose of evaluation is to judge the merit, worth, and value of different evaluands.

Several evaluators have testified that Scriven's suggestion in 1972 about the goal–free model was greeted by stunned disbelief. Completely ignoring objectives was shocking indeed. Scriven's perseverance in making his point, however, seems to have had effects upon the theory and practice of American program evaluation (Guba and Lincoln 1981:16ff.).

What Scriven is reacting fiercely against is the evaluators' obsessive attachment to preordained intervention goals. According to the interpretation of goal–free evaluation provided here, the organizer ought to be results, whether planned or unplanned. By not tilting the evaluation toward stated intervention goals, the evaluator can be more open to the total impact of the evaluand.

For the goal-free evaluator wants to concentrate all her efforts on discovering whatever impacts the intervention has produced. She ought to concentrate on what the evaluand is doing without knowing anything about what it is trying to do. Stated or unstated goals ought to be disregarded. As a matter of fact, Scriven recommends the evaluator to take precautions against locating them and their meaning. For the knowledge of preconceived goals and accompanying arguments may turn into a mental corset impeding her from paying attention to side effects, particularly unanticipated side effects. The major task of the evaluator is to take a global view of the intervention and find out about all the effects. Since she must not identify the goals, she will

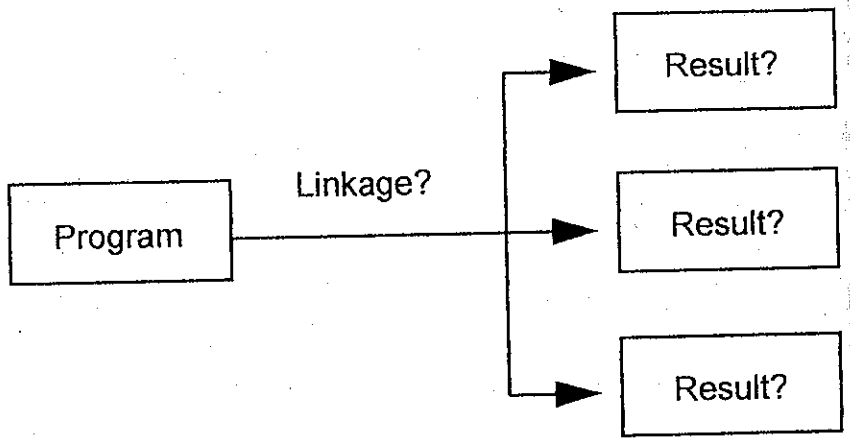
have no basis for distinguishing between intended and unintended effects. She can only ascertain effects.

A skeletal model of goal-free evaluation is displayed in figure 4.6.

Goal–free evaluation is, among other things, an attempt to solve the side–effects problem, the scepter haunting all political planning and goal–based evaluation. It purports to do this by taking a broad view of intervention effects. By consciously avoiding the distinction between main effect and side effects, the model may help the evaluator to trace outcomes that otherwise might have been overlooked. It is argued that goal–based models, whether in the guise of goal–attainment evaluation or side–effects evaluation, tilt the evaluator toward searching for intended effects and overlooking particularly unanticipated and unintended side effects. The goal–free model attempts to remove the negative connotations attached to the discovery of by–products and unanticipated consequences. Scriven maintained that "the whole language of 'side effect' or 'secondary effect' or even 'unanticipated effect' tended to be a put–down of what might well be the crucial achievement, especially in terms of new priorities" (quoted from Patton 1987:36).

In Scriven's own version of the goal–free model, program effects are compared to the needs of the clients, or rather the impacted population. At this point, I have diverted myself from Scriven and ventured

**FIGURE 4.6**
**Goal-Free Evaluation**






a rendition of my own, excluding also needs from the model. In my consciously idiosyncratic reinterpretation, the goal–free model pays attention to neither pre–stated goals nor client needs. In this version, it is less expensive and easier to handle than side–effects models. The problems involved in determining and weighing program goals or consumer needs are inordinately energy demanding and time consuming. In bypassing the rhetorical bog of goals and needs the evaluator will save precious time that can be expended on more pressing tasks.

The goal–free model, in the interpretation presented here, is apparently related to results–oriented management, a notable public sector steering technique, briefly dwelt upon in chapter 3. Actually, *results analysis* would be an appropriate label. Basically, goal–free evaluation suggests that evaluators should just present the facts; the additional steps of applying criteria of merit to the down–to–earth data, integrating the judgments into an overall value and drawing conclusions for future action are best left to the decision–makers. In this way, my reformed notion of goal–free evaluation might find a place in management–by–results contexts.

From the angle of representative democracy, even the reformed goal–free model suffers from one limitation. The goals that it wants to avoid are not haphazard wishes or incidental desires. Goals of interventions adopted by elected political bodies have an especially elated status, since they have been instigated under due procedure by the representatives of the people. It must be in the interest of the citizenry that these goals are taken seriously in evaluations of intervention output and outcomes. This is a perspective that Scriven, even twenty years after he first came up with the model, seems insensitive to. I have been unable to find any allusion whatsoever to this democratic argument against the goal–free model either in Scriven's contributions or in any other literature on the goal–free model.

That my reformed goal–free model apparently provides no criteria by which to judge the merits of outcomes may be regarded as a weakness. Since valuing must be an inherent property of evaluation, it might be argued that the goal–free model is not evaluation at all. It is effects analysis or results analysis, but not evaluation. A semantic dismissal of this type is too narrow, however. The property of being goal–free might well turn into a strength in an evaluation context: the evaluation recipients may do the assessment. The goal–free model would then point toward an institutional settlement of the value crite-

rion problem in public policy evaluation. The solution would be: find out everything about program results but leave their assessment to the pertinent decision makers and power wielders. This would be the philosophy explicitly undergirding my reinterpretation of the goal–free model.

The lack–of–evaluative–criteria observation does not hit Scriven himself, because he has suggested an expedient. "Merit is determined," he says, "by relating program effects to the relevant *needs* of the *impacted* population, rather than to the *goals* of the program (whether the goals of the agency, the citizenry, the legislature, or the manager) for the target (intended) population. It could equally well be called 'needs–based evaluation' by contrast with goal–based (or 'manager–oriented') evaluation" (Scriven 1991:180).

While interesting, Scriven's claims are not easy to handle. Using needs instead of program goals implies entertaining a prescriptive instead of a descriptive view of valuing. But how do evaluators ascertain needs? It seems methodologically much more difficult to elicit needs than to map results and let recipients do the valuing.

A final weakness that the goal–free model interpreted as results–oriented analysis shares with goal–attainment and side–effects evaluation is the omission of costs.

## Comprehensive Evaluations

Comprehensive evaluation models are rooted in the conviction that evaluation should be more extensive than in the all–too narrow goal–attainment model; the process of passing judgments should not be limited to achieved results but include at least implementation, maybe even planning.

It is not the fact that the evaluand[10] is regarded as a constituent of a larger comprehensive whole that is special with comprehensive evaluation models. Specific to comprehensive evaluation is that other component parts of the system than outputs and outcomes are valued, for example, implementation, and feedback.

To clarify the philosophy underlying comprehensive evaluation, I shall use the countenance model, devised by Robert Stake, as a case.[11] Stake identified three stages in the development of an educational program: the antecedent, transaction, and outcome phase. Stake's focus on educational programs has been generalized here to public interventions in general.



Unfortunately, Stake's three phases have no accurate counterparts in public policy or general system language. The antecedent phase consists of events and conditions existing prior to the intervention that may determine, or relate to, its results. In public policy vernacular they may be thought of as "planning" or decision preparation, in systems lingo processes transpiring before inputs are determined. In addition to the program itself, the transaction stage involves processes preceding program delivery as well as program delivery itself. This is covered by "input," "conversion," and "output" in the system model, "public intervention," "administration," and "output" in public policy language. "Outcome" is events and actions immediately succeeding the delivery of the intervention. It might be equated to "outcome" in the general systems scheme and in public policy language.

The major difference between the goal–attainment model and the countenance model is that the former concentrates on the fit between intended and actual results whereas the latter explicitly also involves judging the planning, the decision, and the implementation stages of the intervention.

Roger Kaufman and Susan Thomas (1980:125f.) have provided a pedagogic summary of the countenance model from which the overview in figure 4.7 is slightly adapted and generalized.

Within each of the three major phases of intervention history, a distinction is made between description and judgments. The former category is further divided into intents and observations, the latter into criteria and judgments. This generates the twelve star–marked cells shown in figure 4.7.

In comprehensive evaluation, the first task of the evaluator is to contrast the intended with the actual, the goals with the realities. This confrontation of intents with actual achievements is also typical of the countenance model. Thus, for the antecedent phase, the evaluator would describe the intended preconditions as well as the actual preconditions. What goals were specified concerning antecedent conditions and events? And what actually did take place? For the transaction phase, the intervention as it was planned would be put up against the invention as it was actually carried out. And likewise for the outcome phase: intended impacts would be compared to impacts actually obtained. The three fundamental questions in description would be: Are antecedent events and conditions fulfilled as specified? Is the program carried out as intended? Do the actual impacts conform with those

**FIGURE 4.7**
**Comprehensive Evaluation: The Generic Countenance Model**

**ANTECEDENT (INPUT) PHASE:**
[period of time before the program is adopted and implemented]

**Description:**
- Intents (what goals are specified, what effects are desired)
- Observations (data concerning the activities and events taking place during this phase; description of existing conditions)

**Judgments:**
- Criteria (benchmarks of merit to be used as basis of comparison)
- Judgments (the process of comparing the intents, observations, and criteria)

**TRANSACTIONAL (CONVERSION) PHASE:**
[period of time during which the program is implemented]:

**Description:**
- Intents (the program as it was planned)
- Observations (the actual implementation and delivery of the program)

**Judgments:**
- Criteria (benchmarks of merit to be used as the basis of comparison)
- Judgments (the process of comparing the intents, observations, and criteria)

**OUTCOME PHASE:**
[the period of time immediately following the delivery of the program during which most of the results data are collected]

**Description:**
- Intents (the intended or predicted results of the program)
- Observations (the data gathered concerning actual results at the end of the program)

**Judgments:**
- Criteria (benchmarks of merit to be used as the basis of comparison)
- Judgments (the process of comparing the intents, observations, and criteria)

Source: R. Kaufman and S. Thomas, Evaluation Without Fear, 125f.

desired? From this we also see that goals are used as organizers for the evaluative activities.

The judgments aspect of the countenance model is also divided into two separate concerns: criteria and judgments. Criteria refer to the yardsticks to be used as the basis of comparison. Will the worth of the intervention be judged by comparing it to another intervention? Will the program be compared to a set of absolute standards not directly related to it? Will the program be judged from the point of view of its



own goals? Will the intervention be assessed by referring to something else? Judgments, on the other hand, refer to the overall process of actually comparing the intents, observations, and criteria of merit. It would constitute some sort of overarching balancing calculus.

Kaufman and Thomas maintained that the countenance model forces the evaluator to describe the events, activities, and conditions that exist before, during and after the adoption and implementation of the intervention. This careful and thorough description will provide a wealth of information about both the intentions of those who develop the intervention and on–site observations of what actually occurred during each of the three stages.

Another distinctive feature of the countenance model is its emphasis on the specification of the criteria to be used in the passing of judgments. In the countenance model, the criteria must not remain implicit but be fully explicated.

An additional prominent attribute of the countenance model is the focus on judgments as a major aspect of evaluation, defining the complete act of evaluation as involving both description and judgment (the two countenances of evaluation, as Stake puts it).

The essence of all comprehensive evaluation models is that value achievement is measured in several stages of the system under investigation.

Comprehensive evaluation models are preferable to the goal–attainment model because of their explicit preoccupation with implementation processes between input decisions and actual outputs. This is a strength on two accounts. It allows the possibility of evaluating against so–called procedural goals, that is, procedural fairness, legality, openness in relation to secrecy laws, or popular participation in decision–making processes. Furthermore, the field is open to explain phenomena in the output and outcome stages by factors in the administration processes. This is of crucial importance if the evaluation is to be able to provide an information base for action. If some outcomes did not materialize because some outputs were not made, which is in turn explained by some factor in the implementation stage, then the decision–maker will get information about which something can be done. The outcomes are influenced by outputs, conversion, or inputs. The general environment of the system can also be scanned for clues to influences on the segment of the system under evaluation. The comprehensive evaluation model can be turned into process evaluation.

In other respects, comprehensive evaluation suffers from the same flaws as the goal–attainment model. Its preoccupation with official substantive goals impedes the discovery of side effects, which is a serious fault. Also, the comprehensive model as construed here pays no attention to costs and disregards clients' viewpoints.

Another drawback with the countenance model is its complexity. According to Guba and Lincoln (1981:14), who have touched upon this criticism, evaluation practitioners have found it difficult to comprehend and operationalize the twelve–cell design of the model. From another point of view, the countenance model is too simplistic, because it makes no clear difference between input, conversion, and output, which are all included in the transaction phase.

## Client–Oriented Evaluation

The evaluation models expounded so far have been geared to information wished for by program managers and superior sovereigns. Proceeding from an entirely different point of departure, client–oriented evaluation takes the goals, expectations, concerns, or even needs of the program addressees as its organizing principle and criterion of merit. At the heart of client–oriented evaluation is the question of whether the program satisfies client concerns, desires, or expectations.

Actually, there are two philosophically very divergent points of departure for client–oriented evaluation: client needs and client desires and expectations. Since needs do not necessarily coincide with expressed preferences or demands of the clients, the needs approach involves a prescriptive theory of valuing, that is, the evaluator herself must ascertain which needs are most valid. Establishing needs while retaining a scholarly posture is a thorny issue, which I shall avoid here. Instead, I shall interpret the client–oriented model as based on a descriptive theory of valuing, predicated upon the express desires, expectations, values, assumptions, and objectives of the clients.

There is no common, accepted terminology to denote the targets of public interventions. I have interchangeably referred to them as "addressees," "targets," "clients," and "recipients". As synonyms, I shall use "participants," and, where appropriate, "consumers," and the Janus–like concepts "beneficiaries/maleficiaries."

My extensive experience as an evaluation instructor has taught me to emphasize the difference between program clients (program consumers) and evaluation clients (evaluation consumers). The evaluation client is the person, group, or agency that has commissioned the evaluation of a program, whereas the program client is the intended or actual recipient of the program (Scriven 1991:82). In spite of the fact that expressions like "program users" and "user influence" are very much employed in Scandinavia, I shall avoid it here and reserve the term "users" for evaluation, not program, recipients.



The client–propelled model is justified in two ways. First, it is grounded in political ideologies based on the superiority of the market place over public–sector provision. The belief is that consumer pressures expressed through attitudes toward service provision will lead to the improvement of service delivery and increased consumer satisfaction. In buying a commodity in the store, the consumer pays no attention to producer purposes. Her own assessment of the value of the good is what counts. Client–oriented evaluation is based on the notion that public administration produces goods and services for consumers in the market place. The theory of the market–driven administration is fundamental to client–oriented evaluation.

Second, a democratic, participatory case, is often made for the client–model. According to this argument, the consumer parallel cannot be pushed too far, since the client concept includes a participatory, democratic aspect, which seems to be absent from the consumer concept. The participatory feature suggests that the clients may voice their complaints and desires to the service providers and to some extent influence and take responsibility for service content. The client approach engenders a discursive, reasoning, discussing and, consequently, influencing countenance, which should come to the fore in client–oriented evaluation (Katz and Danet 1973; Goodsell 1981): "the consumer as citizen rather than the consumer as customer" (Jenkins and Gray 1992:296).

The primordial step in the practical application of the client model is to locate the program clients. Since the evaluation normally cannot cover them all, a sample from the target population must be picked. Then the client–driven model points in various directions. The program coverage issue, that is, eventual differences between the intended target population and the actually impacted population, is always important. Is the program covering the whole target population or only parts of it? The crucial activity, however, is to elicit the client's views of the program.

The client model does not tell which program components should be appraised. It allows for a wide variety of assessments. The evaluators may ask clients to pass judgments on some aspect of the service. For instance, clients may be invited to judge program outcome, program output, service availability, service quality, or even service administration. Client–oriented evaluation may also raise the causal issue, that is, the evaluator urges the clients to estimate program impacts. In this case, targets are asked to compare what would have happened had there been no program to what actually has happened with the program in place. Client–oriented evaluation uses the shadow controls design to assessment of program impact (see figure 11.1 and chapter 12).

Another notable feature is that value pluralism is an accepted strand in the client–oriented model. Consumers can disagree in their appraisals. The client approach permits conflicting opinions on public programs and their reception.

It has taken an inordinately long time before the clients were recognized in public administration (Jenkins and Gray 1992:285ff.). The notion of a politics–driven civil service and a parliamentary chain of control—administration as the neutral tool of the elected officials—has exerted a dominant influence over administrative thought. Today, the client–oriented model is employed in numerous evaluative contexts, particularly those concerning public service provision such as urban transit, public utilities, parks and recreation, health services, child care, public housing, and nursing homes for the elderly, where clientele participation is crucial to the operation of the program. The client–oriented model is used to evaluate library services, arts, zoos, and museums. It is a favorite with educators. At universities, students are routinely requested to share their opinions of courses, reading lists, and lectures. They are asked to rate their teachers' abilities to organize the course contents, to stimulate and promote altercation and discussion, to stir student interest, motivation, and critical thinking, and to show concern and enthusiasm for the students. At American universities, these evaluations are occasionally used to rank faculty and courses from a student perspective.

Here are some examples of market research done in the Swedish state sector: the National Audit Bureau (RRV) wanted to know the taxpayers' views on the service level, staff competence, and performance of the taxation authorities; the National Labor Market Board asked how job seekers perceived the services of the local employment offices; the National Board of Public Building wanted to know its clients' (tenants') opinions of how the board manages the provision of premises; the Ministry of Health and Social Affairs asked a large number of parents for their opinions on child care; the National Environmental Protection Agency surveyed the credibility of various environmental agencies in public opinion; and the Patent and Registration Office investigated how inventors perceived queuing time.



In my view, the client–propelled model may complement the previously presented approaches, since it poses another problem for consideration. It can make important contributions to evaluation, but should not be allowed to replace the other models. The exclamation of the Swiss hotel proprietor Cesar Ritz, "*Le client n'a jamais tort*" ("the customer is never wrong"), cannot be the lodestar of all evaluation. The requirement that the civil service must be responsive to client concerns is sound, but only within limits. It can never take precedence over the requisite that the front–line operators should follow the directives of their hierarchical administrative superiors, and indirectly political bodies like Parliament, and the Municipal Council, and ultimately, the citizens. Tax authorities must answer promptly and correctly to inquiries from the general public, but they cannot lower taxpayers' taxes or make haphazard exemptions. As a rule of thumb, I would argue that clients' criteria of merit ought to play a more prominent role in public service assessment than in the assessment of regulatory regimes. But evaluators must be aware of the tendency of the clientele to exaggerate complaints in order to get more service. Greater client involvement in evaluation may surrender power to groups with vested and narrow interests (Jenkins and Gray 1992:296).

## The Stakeholder Approach

The organizing principle of the stakeholder model is the concerns and issues of the people who have an interest in or are affected by the intervention. This is quite different from using prefixed objectives as merit criteria, as in goal–attainment evaluation or even side–effects evaluation. Stakeholder evaluation, however, does resemble the client–oriented model, the major difference being one of scope: while the client–driven model is concerned with one group of affected interests, the stakeholder model is geared to all of them.

Stakeholder–based evaluation starts by mapping out the major groups who are involved or have an interest in the emergence, execution, and results of the program. The evaluator identifies the people who hammered out the program, who initiated and funded it, and particularly those who are charged with its implementation. She identifies senior managers, middle managers, lower managers, and front–line operators, who actually deliver program output. She singles out the intervention's primary target group and those who know they have a stake in the program but prefer to keep a low profile. And she identifies those who are unaware of the stake they hold.

An overview of conceivable stakeholders in public interventions (Guba and Lincoln 1989:40f.; Riecken and Boruch 1974:203ff.; Rossi and Freeman 1989:422ff.; Weiss 1972a:18) is presented in figure 4.8.

The fourteen groups listed under "Program Stakeholders" in figure 4.8 represent the major parties interested in or affected by a particular government intervention. The list is compiled with some national reform—decided by parliament—in mind. Naturally, a listing of stakeholders in a smaller agency program or some program at the municipal level would look differently. Actually, the stakeholder model provides no clear answer to the question of who the stakeholders are. It is open–ended in this respect. In figure 4.8 I have also included three potential stakeholders, who may not be interested in the program as such but in its evaluation.

The fourteen major stakeholding audiences are all conceivable participants in an evaluation of the intervention. I shall also assume that all the stakeholders are immersed in the evaluation. The stakeholder model will be interpreted as an holistic approach, directed at the whole spectrum of affected groups and their organized emissaries.

At this point, I shall distinguish between the Swedish and the North American stakeholder model. In the Swedish model, the stakeholders actually execute the evaluation and take responsibility for its findings. In the North American model, by contrast, the evaluation is not performed by the stakeholders but by one particular evaluator or a team of evaluators, who may consist of consultants, employees of the agency in charge, or university researchers. The evaluation team should then unravel and collect stakeholder concerns, expectations, and interests and bring them to bear on the evaluation. While the stakeholders are contacted, provide the problems to be investigated and the criteria and standards to be used as instruments of appraisal, the evaluators—not the stakeholders—carry the full responsibility for the final results.



**FIGURE 4.8**
**Stakeholders in Public Interventions and Their Evaluation**

| | |
|---|---|
| **PROGRAM STAKEHOLDERS:** | |
| **Citizenry:** | Citizens who elect representatives to decision–making bodies at all levels in the political system. |
| **Decision–makers:** | Political officials responsible for deciding whether an the evaluated intervention is to be instituted, continued, discontinued, expanded, or curtailed. |
| **Political Opposition:** | Political opponents of the intervention. |
| **National Agency Managers:** | Senior managers in charge of the intervention at the national level. |
| **Program Directors:** | Middle managers in the national agency directly in charge of the intervention. |
| **Regional Agency Managers:** | Administrative unit at the intermediate level responsible for some aspect of the implementation of the intervention (e.g., regional inspectors). |
| **Private Intermediaries:** | Non–governmental institutions charged with some responsibility for the implementation. |
| **Local Agency:** | The lowest–level administrative unit responsible for the delivery of the intervention (e.g., a local bureau of a national regulatory agency). |
| **Street–level Bureaucrats:** | Front–line operators responsible for program delivery, and who directly communicate with the addressees either face–to–face or via telephone, mail etc. |
| **Clients:** | Individuals or collectivities like households, businesses, organizations, municipalities or other units who are the addressees of the public intervention under evaluation. |
| **Neighboring Agencies:** | Governmental units in charge of interventions related to the one under evaluation. |
| **Program Competitors:** | Organizations or groups who compete for the available resources. |
| **Contextual Stakeholders:** | Organizations, groups, individuals, and other units in the immediate environment of an intervention. |
| **Research Community:** | Scholars specializing in the substantive issues which the intervention is concerned about. |
| **EVALUATION STAKEHOLDERS:** | |
| **Evaluators:** | Persons responsible for the design, conduct, and findings of the evaluation. |
| **Evaluation Sponsors:** | Organizations that initiate and fund the evaluations. |
| **Evaluation Community:** | Other evaluators, and teachers of evaluation methodology, who read and judge evaluations for their technical and substantive quality. |

Once the major interested parties are discerned, the North American stakeholder model might function in various ways. In their book *Effective Evaluation,* Egon Guba and Yvonna Lincoln (1981:33ff.; also 1989:50ff.) proposed that stakeholder "concerns" and "issues" should be the starting point. A "concern" is "any matter of interest or importance to one or more parties." It may be something that threatens them, something they fear might lead to undesirable consequences for them, or something they are anxious to substantiate. "Virtually any claim, doubt, fear, anticipated difficulty, and the like expressed by anyone with a legitimate basis for making such a representation could be entertained as a concern." An issue, on the other hand, "is any statement, proposition, or focus that allows for the presentation of different points of view; any proposition about which reasonable persons may disagree; or any point of contention."

To build a fitting general design for the study, advocates of the North American stakeholder model nurture a strong penchant for qualitative methodology. The key word is *interactive* search procedure. The evaluator must talk to the stakeholders to elicit their narrative histories and observational data, which in turn should be allowed to affect the evaluator's next step in the search procedure. After a while, she might discover both the purported and the genuine aims of the program, and what concerns various stakeholders nurture regarding evaluand and evaluation. With time, the evaluator gets more involved and can start to determine what concerns and issues should be included in the study. Only then can she take a stand on what the outline of the evaluation should be.

It is typical for the stakeholder model that the evaluator is permitted to search rather extensively for the crucial problems. The goal–attainment and the comprehensive evaluation models point out much more specifically which problems to raise and which questions to ask in the investigation. The North American stakeholder approach is remarkably open in this respect, empty if you like. The idea is that the evaluator must be responsive to the concerns and issues of the affected people and let these govern the next step in the investigatory enterprise. Through interactive communication she is supposed to find out which concerns and issues are to be taken seriously and probed more deeply. The evaluation design will be gradually determined. Stakeholder evaluation is *responsive evaluation* (Stake 1975; Shadish, Jr. et al. 1991:275ff.).



To elicit the final data on stakeholder concerns and issues, advocates of the North American stakeholder model seem to prefer observational and interrogative methods to documentary methods. In many cases, stakeholder evaluators endorse direct or even participant observation. In–depth interviewing is another favored technique. After data are amassed and processed, the reporting of results, which might vary from one stakeholder to another, will commence. The key word seems to be "portrayals," that is, information–rich characterizations using pictures, anecdotes, thick descriptions, and quotes. The comprehensive holistic view mediated through a portrait is important. The evaluator's own value criteria should play no role at all since the selection of measuring rods as well as the measurement proper is supposed to be the task of the stakeholders. Normally, several criteria of merit, standards of performance on these criteria and several comprehensive assessments will be included. In practice, the stakeholder model will become downright pluralist. Finally, it is decided whether the results should be collected into a written report—something that Guba and Lincoln (1981:25f., 39ff.) consider by no means necessary.

The stakeholder model, it is argued, has numerous *advantages.* A knowledge argument, a utilization argument and a goal–management argument may be adduced in its favor.

According to the *knowledge argument,* it would be foolish of the evaluator to avoid the program knowledge, which those involved indubitably have. Stakeholders nurture convictions about side effects, implementation barriers, and outright cheating, which may provide ideas for continued research. Consequently, it is very easy to agree with the recommendation that almost every evaluation ought to begin with the determination of relevant actors and rounds of interviewing.

The knowledge argument, however, is usually accorded less weight than the *utilization argument* in support of stakeholder evaluation. In North American evaluation debate, stakeholder involvement is regarded mainly as a vehicle for increasing the use of the evaluation in upcoming decisions. Findings from traditional goal–attainment evaluations have little impact. The reports are buried in desk drawers and heaped on the bookshelves of the public authorities involved, unread and forgotten. Although the will is there, decision-makers do not avail themselves of the information base in the evaluations to make more rational decisions. The business of evaluation suffers from a malign "desk-drawer syndrome."

This seemingly wildly irrational behavior has puzzled and exercised researchers. Why do decision-makers behave in such an unwise fashion? The allegedly major explanation is that evaluators work in splendid isolation, with too little communication with prospective users. They tend to investigate problems on which there is no pressing demand for information. The stakeholder approach increases the chances that questions of genuine interest to concerned parties will be addressed. It brings to light information that meets the real requirements of the different stakeholders, thereby enhancing the probability that the results actually will be put to use. This is a significant point, since goal-based evaluations often wind up being ignored, without influencing the course of future events.

The *goal–management argument* views the stakeholder model as a strategy to handle situations in which there are no written premeditated goals or where the stated goals are notoriously unclear and difficult to trade off against each other according to accepted scholarly canons. We have already seen that not even the goal–attainment model can fully master the criterion issue, particularly the multiple goals issue so typical of public policies. The stakeholder model, however, provides a practical solution to the problem of eliciting concerns and issues where no written objectives exist or of simultaneously managing several contradictory goals.

Individual implementers may have no clear directives from above concerning what they should do, because with almost no exception, higher–level policy goals are general, contradictory, and muddy. Most importantly, no trade–offs are provided between the several goals that are claimed as lode stars for the policies. This makes it impossible to use given intervention goals as objective yardsticks by means of which the global merits of the policies may be judged. One way out is to regard the choice of overall criteria and standards as an inherently political task. The evaluators should describe the value positions of many different stakeholders, without contesting any of them. The fact that stakeholders disagree is normal and natural in politics, and a sign of health in free nations. Findings should be presented separately on each goal. Actually, it might be advisable to avoid doing a single large evaluation study and conduct several smaller studies instead, geared toward specific stakeholding constituencies. In no case should the evaluator render summative judgments. Overall summation is best left to the stakeholders. The stakeholder model uses a descriptive theory of valuing.



There are also obvious *drawbacks* with the stakeholder model. In most versions, it ignores program costs. In addition, it is inordinately impractical and resource demanding, since every stakeholding constituency must be contacted and nurtured.

The stakeholder model is also fuzzy. It provides no authoritative answer to the question of who the stakeholders are. The range of stakeholders must be decided on a case–by–case basis. There is a need to supplement the stakeholder approach with a political or administrative theory concerning the selection of affected groups. Furthermore, all of the interests concerned, however selected, are treated as equals by the stakeholder model. But in a democratic, constitutional system, elected politicians must carry more weight than administrators or experts on policy instruments and the substantive matters under consideration, just to pick a few. The stakeholder model does not prioritize among the diverse stakeholding audiences. The model is grounded in a modest equilibrium philosophy. There is a risk that the best organized and most committed stakeholders are consulted while vaguely concerned groups are omitted. Even in this regard, the stakeholder approach is in need of a more developed philosophical base.

The most serious objection, however, is the risk that the stakeholder model will embrace a pragmatic theory of truth. Truth can turn into a matter of usefulness, utility, or acceptability to stakeholders. Stakeholders often entertain highly politicized views of program effects. Opponents ascribe everything negative occurring after the program as caused by the program and everything positive as caused by something else. Supporters hold the opposite view. Facts are essentially contested. In these situations the various parties will accept only those findings that fit into their preformed opinions. Too much stress on utility and usefulness can turn evaluators into substituting well–corroborated information to less validated findings that are useful and used by somebody. In my opinion, evaluators must always be on guard against pragmatic leanings. If they start to deliver useful instead of true information, they act as political animals, not as scholars defiantly searching for truth.

In conclusion, while quite controversial, the stakeholder model carries some important merits. The utilization and the goal–management arguments speak in its favor. Another strong reason for it is the knowledge argument. However, stakeholder protagonists, who want to claim they are researchers, must beware of the risks of substituting usefulness to truth.

## A Swedish Stakeholder Model: The Ad Hoc Policy Commissions

The definitional difference between the North American and the Swedish stakeholder model is that in the latter the stakeholders perform the evaluation and take responsibility for its results whereas in the former the stakeholders are only consulted and allowed to influence, design, value criteria, and other tenets of the evaluation but the evaluation itself is carried out by some independent evaluators. In the Swedish system of ad hoc policy commissions, evaluation is performed in five partly consecutive, partly parallel, processes: commission work, consultation procedure, public policy debate, government decision, parliamentary decision.

In the Swedish political system, a considerable source of information on past performance and alternatives for future policy action is produced by ad hoc investigative bodies. A highly visible subset of such bodies are the *ad hoc public policy commissions,* in Swedish called *offentlig utredning* or *kommitté,* on rare occasions also *kommission,* or *delegation.*

While the formal initiative may be taken in Parliament, it is always the government *in corpore* that summons ad hoc policy commissions and specifies the problems to be investigated. The written directives, which are made public, indicate what a particular commission should do, determine the economic frames, and designate a completion date for the work. The government also appoints the chairman and other members, experts, and participants.[12]

The policy commissions usually consist of full members, political appointees (*sakkunniga*), experts, and a secretariat. The latter is composed of one full–time administrator, who functions as secretary, and a few part–time secretaries. The format of the commissions and the background of their participants will vary according to the subject under investigation. For highly sensitive matters, such as defence or constitutional reform, the entire commission is frequently made up of full members who are parliamentarians, from parties in office as well as from the opposition. For matters moderately political, full members of all camps in Parliament will be appointed along with representatives of other stakeholding constituencies such as ministries (which are kept amazingly small in the Swedish polity because agencies are rather independent of ministries), national implementing agencies, the Association of Local Authorities, the trade union movement, and organized business. In most cases, people from academia are hired as experts and fact finders. In addition, the commissions themselves are allowed to contract out specific assignments to scholars, who then do their business as independent researchers.



A typical feature of the Swedish commission approach is that *ex post* evaluation is married to prospective policy analysis. Actually, investigations have traditionally focused on alternatives for future action rather than impacts of past policies. Besides an historical account of how the pertinent evaluands as they are codified in the books have evolved over the last decades, the typical commission report includes data on the severity and scope of the problem, substantial prospective assessments of different options or the most promising option for future policy action, and a sometimes clear recommendation of policies. Yet systematic information on the implementation, output, or outcomes of earlier interventions is usually also provided, although to a lesser extent. In the last two decades, however, more emphasis has been placed on the provision of evaluative information. Swedish policy–makers have learnt the same lessons as their counterparts in other countries: implementation according to plan as well as achievement of intended outcomes are problematic. Now, some commissions even have evaluation as their major task (SOU 1981:40–41; the Price Controls Commission).

Sometimes the commissions also perform before–the–fact implementation analyses. They attempt to scrutinize the staff, organizational, and managerial capabilities to determine the degree to which the proposed policy alternative(s) can be specified and brought to street–level execution in a particular bureaucratic setting. This feature is particularly interesting against the backdrop of clamors for *ex ante* implementation analysis as the missing link in public policy analysis.

The conclusive tracts are invariably signed by all full members of the commission. This means that the stakeholders shoulder the complete responsibility for the content, and particularly for the recommended actions. However, members have a right and a duty to register dissenting opinions in so–called reservations (*reservationer*), which are added to the tract. Also, experts may submit dissenting opinions in so–called special utterances (*särskilda yttranden*), which are appended to the report.

The one–, two–, or even three–volume studies (*betänkanden*) are always published, which is a testimony to the comparative openness of

the Swedish state policy–making system with its constitutionally regulated presumption against secrecy, the so–called Principle of Publicity. The reports are issued in a series of volumes referred to as Government Official Investigations (*Statens Offentliga Utredningar,* abbreviated SOU). All volumes carry ISBN and ISSN numbers. The tract is primarily addressed to the government. Yet it is also readily accessible to all conceivable stakeholders, citizens and mass media included, since it is available in every municipal and other major library throughout the country. Certainly, the SOU–Series provides worthwhile sources of policy–relevant research into most aspects of Swedish social life.

Commission work constitutes the first stage of policy preparation. In a second stage, the ad hoc commissions are coupled to another well–entrenched Swedish structure of stakeholder evaluation, the *comprehensive consultation procedure*—called *remissväsendet, the remiss procedure*—which is adopted after a commission's proposals are made public. The government sends the published commission studies to numerous interested parties for written reactions. These parties usually include government agencies at the national and regional level, county and municipal governments, universities, institutes of technology, various kinds of interest organizations, trade associations, and professional organizations. The parties must provide their answers within the time limit set by the government. Like the commission reports the consultation answers are also made public.

The remiss procedure can be regarded as metaevaluation in the sense of auditing of a final evaluation. To the extent that the commission reports contain evaluative results, the stakeholders are asked to evaluate an evaluation.

The metaevaluations of the commission reports also take other forms. Since both the commission reports and the consultation answers are open to the public, they might receive wide coverage in the mass media, including front page articles and radio and television news reports. In addition, they may trigger extensive and prolonged public policy debate. Participants in these debates usually include daily newspapers of all political shades from all corners of the ablong country, the national and local television and radio channels, the media of the trade union movement, the business communities, the nonprofit associations. In addition, independent researchers, politicians, representatives of interest organizations, and numerous other categories contribute articles to the newspapers and reveal their opinions in interviews.



The *public policy debate* can be regarded as the third component of the policy formation and evaluation in Sweden. Rather than a separate stage, it should be viewed as a parallel process to the consultation procedure and later stages.

On the basis of the commission proposals, consultation answers, the mass media debate, and probably also other information, a draft proposal (*proposition)* is hammered out in the pertinent ministry, usually in continued close contact with some interested parties. While the elaboration of the *draft bill by the government* may be considered the fourth process in decision preparation, the *parliamentary processing of the bill* may be considered the fifth. When the draft bill is finally submitted to parliament, it is almost routinely approved without major changes. The formal supreme decision–making body, the legislative assembly, exerts minor influence on real decisions. All in all, decision preparation is a very long process with a multitude of actors and widely divergent interests involved where compromises are struck and the issues are decided well before the issues formally enters the parliamentary or even the governmental arena.

Exactly how evaluation–based information is used in the deliberations of the commissions, of the interested consulted parties, in the media debate, in government, and in the diet is difficult to pinpoint. Occasionally, it probably carries some weight in substantive deliberations, albeit far from decisive. In retrospect, however, stakeholders often claim that they make their political judgments, negotiate and strike their political compromises on the basis of the information amassed by the experts. The same goes for the government. Presumably, evaluation is somewhat utilized to render legitimacy to the proposals suggested. Perhaps it also gives some substantive guidance.

Indisputably, the Swedish ad hoc policy commissions are cases of the stakeholder model in action. Yet, the practice deviates in several crucial respects from the stakeholder model expounded in North American literature.

To a much larger extent than the North American stakeholder model, the ad hoc commission institution reflects the view that evaluation is primarily a *political enterprise,* not research work. While a large spectrum of affected interests are consulted and carefully listened to, in the North American stakeholder model it is still the evaluator—that is, a researcher—who is in charge, decides which stakeholder problems to penetrate, how to conduct the evaluation, and also assumes responsi-

bility for the final report and other outcome of the evaluation. The Swedish stakeholder model is more openly political. First, the government establishes the commission, selects the participants, determines the issue agenda, and allocates the funds. By contrast with the North American model, the Swedish stakeholders are full members of the investigation team, execute or at least oversee the execution of the evaluation, and assume complete responsibility for the findings as well as the recommended options for action. Yet, they are mandated to ponder the issues and work within the time and resource constraints specified by the government. Third, performed within a political framework, the commission work is specifically geared to a particular recipient–cum–decision–maker, the national government, and designed to meet a specific pending decision situation before a set deadline.

Another difference concerns the *view of summary statements*. In the North American stakeholder model, there is a tendency toward separate reporting to each stakeholding constituency; assignment of a single global value to an intervention is regarded with suspicion. This is somewhat alien to Swedish official thinking. The philosophy embraced by the commissions strongly suggests that a unanimous final report is something valuable to strive for. Evaluative data should serve the whole spectrum of political interests, not just particular stakeholder concerns, and enable full members to reach unanimous solutions. While minorities have a right and duty to make reservations, the alternative preferred by the reigning political culture is unity.

The accompanying *media attention* seems to be much stronger in the Swedish case, partly because the process is institutionalized since at least a century, and maybe also because of the openness as a consequence of the Principle of Publicity.

Another dissimilarity is that the commission institution is used not only to come up with policy proposals but to effectuate *compromise* on policy implementation as well. Hydroelectric power policy is an excellent case in point. The parties in Parliament already in 1975 set a very clear quantitative goal for hydroelectric construction. But the political parties and other concerned stakeholders have not been able to reach agreement on which particular water courses and river falls should be dammed and exploited, despite numerous attempts with ad hoc policy commissions and consultation procedures. A more clear–cut example of the saying that "implementation is not neutral rule application but the continuation of politics with other means" is difficult to imagine.



A basic similarity between the two models is that results should not be judged against one set of criteria of merit—stated policy goals—but a wide variety of yardsticks, represented by the views of the numerous stakeholding constituencies.

Another commonality concerns the role of research. While the research feature is stronger in the North American approach, it is also in the Swedish case thought that program evaluation, while basically political, ought to be fused with scholarship. The research being executed or collected is discussed in the commissions and the research papers are often presented as annexes to the main report and under the name of each researcher. Parts of the research findings are also included in the major commission text backed by the majority of the full members of the commission.

The remiss procedure bears stronger resemblance to the North American stakeholder model than the commission preceding it, above all because now the stakeholding constituencies are asked individually to react to the commission investigation without being members of any formal appointed body.

The approach taken by the ad hoc public policy commissions is justified somewhat differently compared to the North American stakeholder model. The argument that stakeholder involvement will increase evaluation use is never heard of in Swedish policy communities. Instead, the basic rationale is very Swedish in the sense that the commission model is intended to promote compromises, agreement, support, and forestall bitter political struggle. Commissions are consensus–building mechanisms rather than rational problem–solving institutions. They are vehicles for shaping agreement on the results of earlier efforts, and most importantly, proposals for future action. The appointment of an ad hoc commission marks a political attempt on behalf of the incumbent party or parties to reach accord with the parliamentary opposition, the strong interest organizations, and the public interest associations before the government drafts a bill, and well before the draft bill is submitted to Parliament. Sometimes, commissions are also used to promote accommodations among the dissenting governing parties themselves. Consensus building and the rendition of legitimacy to fundamental decisions are major arguments in favor of the commissions. The underlying assumption is that people will have more confidence in a policy in whose development they were consulted, although in the final analysis it does not agree with their particular preferences.

Finally, a democratic, participatory case can be made for both models. Through the use of the stakeholder approach, affected interests can participate and influence the final outcome, which is of democratic value by itself.

Occasionally, the commission institution is sharply castigated. Commissions are time consuming. The time–consuming argument carries weight because it may take years before the final report(s) are presented to the public. Another critique suggests that commissions tend to erode the authority of the parliament. Once the process reaches the elected officials in parliament, the issue is politically resolved. The role of the parliament will be to rubber stamp what has been decided elsewhere. Also, this stricture carries some weight. A third objection concerns the well–established fact that commissions are also used strategically, as a dilatory procedure, to entomb difficult issues and have them dropped from the public agenda.

"Virtually every important piece of legislation is prepared through the work of specially appointed governmental commissions," Rune Premfors (1983:623) wrote in an article at the beginning of the 1980s. While correct in the past, this statement no longer reveals the whole truth. The sheer number of policy commissions has declined in recent years. In the middle of the 1960s the amount of commissions increased from approximately 250 to 300. During the Center–to–Right coalition period in 1976 through 1982 a peak was reached with 422 commissions in 1980. In the fall of 1983, 206 commissions were at work; in 1987, 195. The average lifetime of a commission used to be three to four years. Lately, efforts have been made to speed up the investigative process. Now, a commission may work only for two years or less. The reports are also supposed to be reduced from 300 to less than 100 pages.

The commissions' influence on state policymaking is probably also less than before. There are moments when time is of the essence and these situations require more rapid modes of decision preparation. For this reason, and maybe also for a political propensity to avoid public debate on sensitive matters, several major settlements are preceded by no public inquiries at all. A blatant case was the decision to apply for membership of the European Community and with several decisions on cross–bloc crisis packages in the fall of 1992. Increasingly, ad hoc commissions of inquiry are also replaced by departmental investigations, commissioned by a minister—not the cabinet as a whole—and



produced entirely by officials from within the pertinent department. These studies are published, if at all, in the less accessible Ds–series. Furthermore, it also seems that agency investigations conducted within the appropriate state agency play an increasingly more consequential role.

## Economic Models: Productivity

Common to all effectiveness models is their negligence of costs. However resource guzzling a program has been, effectiveness evaluators concentrate on substantive results and disregard costs. Attention to costs, on the other hand, is a typical feature of the economic models of public policy and program evaluation.

Often, private business is held up as an ideal to emulate in the public sector. In private business, profit is the criterion of merit and maximizing profit is the standard of success. In analogy to this, productivity maximization ought to be the standard of good performance for public interventions. Productivity seems to be an unusually clear concept; it is defined as the ratio of outputs to inputs. Or, to quote Wholey and Newcomer (1989:144): "Simply stated, *productivity* is the relationship between output of products and services and input of resources: output divided by input."

Productivity can be expressed through the mathematical algorithm shown in figure 4.9.

Evidently, the ratio formula in figure 4.9 can be operationalized in many ways. Let me provide a Finnish example from the library community. In computing the productivity of Finnish municipal libraries the following measure has been used (Sjöblom 1991:12ff.):

$$\frac{\text{number of books borrowed}}{\text{costs in Finnish Marks}} = \text{cost productivity.}$$

As an alternative to cost productivity people have resorted to work productivity, which can be illustrated by the expression:

$$\frac{\text{number of books borrowed}}{\text{number of hours worked}} = \text{work productivity.}$$

**FIGURE 4.9**
**Productivity**

$$\text{Productivity} = \frac{\text{output}}{\text{input}}$$

The difference is that costs in the former case are indicated in monetary terms, in the latter case in the number of hours worked, that is, as physical entities. It should be emphasized, perhaps, that costs can be computed in both ways in productivity measurement. The time unit used could be the fiscal year, the calendar year, or even a monthly period.

Other possibilities would be the ratio of library holdings (number of books kept) to costs, the number of inhabitants in the municipality to costs, or the number of borrowers to costs.

To say something worthwhile about the actual productivity of a public agency, a reference case is needed. That is, evaluators need performance standards on the productivity criterion to tell what high or low productivity means. Several standards are used: comparison with past performance, with similar institutions in the same country, similar institutions in other countries, goals of the political bodies, client goals or stakeholder goals. An overview will be given in chapter 14.

Productivity as a measure of public sector activities carries some technical advantages. Sometimes, costs are not terribly difficult to calculate since they reach the agencies in terms of monetarized funds. However, at other times it may be hard to trace the relevant costs because the funds allocated and spent are often indicated as lump sums on the books. Outputs, on the other hand, may be exacting to catch and compute, even though productivity only presupposes that they are indicated in physical, not monetary, terms.

Let me return to the Finnish municipal libraries to illustrate the difficulties of finding valid output indicators. Is the number of borrowed books really a relevant and exhaustive output measure? Admittedly, it is relevant. To provide the public with an opportunity to borrow books is reasonably the most important task of a public library. But it is certainly not exhaustive. A Finnish report concludes that only 30–50 percent of the library clients borrow books. The other



patrons visit the library to read newspapers, magazines, and journals, but this does not show in the borrowing statistics. They frequent the reference library to use dictionaries and encyclopedias or the music department to listen to records, tapes, and discs. Libraries also provide some information services (Sjöblom 1991:18ff.). In the Finnish investigation I am referring to, these problems were noticed and discussed before the researchers decided to choose the number of borrowed books as the indicator of library output. But this does not turn the chosen measure into a fully exhaustive one.

Another intricacy involved in productivity evaluation is that qualities often are overlooked. Books differ in quality. How can that be measured in productivity assessment?

There are other criticisms of productivity as a measure of the virtue of public policies. The most important suggests that productivity is an internal measure, which does not apprehend what we really want to disentangle, namely, the results that the outputs have produced with the end receivers or in society at large, the value of these results, and if the benefits are worth the costs. In the library example, borrowed books are not necessarily significant themselves; people may charge out books from the library, place them in a heap on the desk at home, and after some weeks return them unopened. More important is the reading of the borrowed books. But what really matters are the borrowers' gains from their reading. The gains may be recreational or educational. What the cost–conscious, education–oriented library evaluator really wants to grasp may be:

$$\frac{\text{value of education through books borrowed}}{\text{costs}}$$

But if so, then she has left productivity measurement and entered the field of efficiency evaluation.

Proponents of productivity as a criterion against which to evaluate public policies, programs, and services cannot escape the fact that productivity as a yardstick of output is not an ideal measuring rod for assessing the worth and merit of public sector activities. The public institution may do wrong things, that is, the outputs may not produce the desired outcome.

## Economic Models: Efficiency

The second major economic model to be used in modern political and administrative evaluation is the efficiency model. Efficiency can be measured in two ways, as cost—benefit or as cost—effectiveness. "Efficiency assessments (cost—benefit and cost—effectiveness analyses) provide a frame of reference for relating costs to program results," write Rossi and Freeman in their widely used textbook *Evaluation: A Systematic Approach*. "In cost—benefit analyses, both program inputs and outcomes are measured in monetary terms; in cost—effectiveness analyses, inputs are estimated in monetary terms and outcomes in terms of actual impact" (1989:375).

If measured in a cost–benefit analysis, efficiency can be expressed as the ratio of the monetarized value of the outcomes produced by the program to the monetarized costs. If equalized to what is measured in a cost–effectiveness analysis, efficiency pays heed to monetarized costs as in cost—efficiency analysis, but the value of the effects is indicated in physical terms only.

This is expressed in the simple algorithms in figure 4.10.

"Program effects" in figure 4.10 indicate consequences produced by the program. Effects are not identical with all occurrences in the target area after the instigation of the program. These occurrences may have been caused by something other than the program. What we are looking for in efficiency analysis are effects produced by the program, and nothing else. On this account, efficiency assessment (both cost—effectiveness and cost–benefit) uses the same measure as effectiveness analysis. The major difference is that efficiency takes costs into account, which is not the case in effectiveness analysis.

Like all other evaluation models, productivity and efficiency measurement provide partial perspectives. They overlook other requirements normally demanded from public sector activities in contemporary democracies. Examples of such value criteria are legal equity, procedural fairness, representativeness, participatory values, and publicity rules. No productivity and efficiency study or any other evaluation for that matter, however pretentious from a scientific and scholarly point of view, can explain in an objective fashion how a balance should be struck between these values and productivity/efficiency. The trade–off can only be made through public debate, opinion formation, compromise, and eventual majority decisions, that is, through politics.



**FIGURE 4.10**
**Efficiency as Cost–Benefit and Cost–Effectiveness**

$$\text{Efficiency (cost–benefit)} = \frac{\text{value of program effects (in SEK, US\$)}}{\text{costs (in SEK, US\$, etc.)}}$$

$$\text{Efficiency (cost–effectiveness)} = \frac{\text{program effects in physical terms}}{\text{costs (in SEK, US\$ etc.)}}$$

## Professional Models: Peer Review

Professional models imply that members of a profession are entrusted to evaluate other members' performances with respect to the profession's own criteria of merit and quality standards of performance. The evaluation is conducted by a collegium, which by definition is an assembly of professional equals, so that lawyers evaluate lawyers, professors evaluate professors, surgeons other surgeons, and so on.

The peer review, the most celebrated professional model, has mostly referred to a screening procedure for the selection of contributions to scientific journals. Submitted articles are subjected to peer review to decide whether they ought to be accepted for publication. Peer review is also used to offer guidance to research foundations concerning which projects should get funding. Research proposals are submitted to a group of respected colleagues for screening. On top of that, peer review has also been used to investigate and judge supposed transgressions of rules of ethical conduct and inherited research practices.

However, in this context peer review is a procedure for retrospective assessment of implementation, outputs, and outcomes of public policies. Peer review is particularly aimed at performing an overall quality judgment of the evaluand.

The peer review model is constantly used in *research evaluation*. Renowned scientists of a particular field are assigned to assess the quality and relevance of a research project, a research program, or a university department. And in Sweden, recently, research institutes (Wittrock et al. 1985) as well as disciplines as a whole across the country have been evaluated using international peer groups (Öhman

and Öhngren 1991; Engwall 1992). Scientists and their achievements are evaluated by their respected colleagues and equals. Peer review of research usually is—and ought to be—interactive; usually the procedure starts with self–evaluations by the evaluatees; on the basis of these and additional material like documentary evidence and site visits evaluators pass their preliminary judgements; then evaluatees are given opportunity to comment on the evaluators' reports before they are finalized; all in all, the evaluators listen to the evaluatees and solicits their opinions.

Peer review is an institutional model of evaluation. It does not indicate the substantive questions to be asked but how the evaluation is to be organized. The model only tells you that evaluations ought to be carried out by equals. The task of selecting and applying criteria of merit and standards of good performance is left to the professionals themselves.

We might wonder what peer review has to do with public sector evaluation. The model seems miles away from the arena of policymaking and program enactment. The answer is embodied in the principle of the profession–driven public sector. In some areas of public life, goals are so complex and techniques so difficult, that political officials have found it wise to leave the shaping and debating of them to well–educated professionals. Architects, judges, professors, doctors, veterinarians, and engineers would be cases in point. Hence, it is also considered natural to delegate *ex post* evaluation to the professions. But since these professionals work in the public sector, peer review must be regarded as an evaluation model on a par with the other models used in public life.

The evaluation of government–funded basic science and applied research can be thought of as a review of the state's research policies. Research is performed neither by Weberian hierarchies, nor by interest organizations or elected politicians. It is carried out by professional scientists and scholars. Hence, the quality of their work is frequently evaluated in collegiate forms. In research evaluation, experts often work on assignment for somebody, for instance, a research council.

To provide an example, the evaluations sponsored by the Swedish Council of Building Research at the end of the 1980s were assigned to take special interest in:

a. the relevance of problem selection and design of analysis;



b. the suitability of the methods of analysis;
c. the tenability and validity of arguments and conclusions;
d. the work in relation to the discourse in the pertinent area;
e. the practical applicability of research findings;
f. the worth of the research community shaped by the project;
g. agreement between original intentions and findings reached (Nilstun 1988:51).

After the directives had been written, appropriate experts were approached. Preferably, the experts should have more specialist knowledge in the field than the colleagues whose research is to be evaluated. They should also be independent; for instance, they must not have carried out research work in the area under scrutiny in cooperation with the people to be evaluated. There is an important difference between those peer reviews where the evaluatees has suggested and agreed in advance on the choice of experts, and reviews where they have no say in peer selection.

After the expert group, the "peer collegium" as it were, had been chosen, the upcoming work was organized. Invariably, the reviewers and the reviewees interacted with each other during the reviewing process. To the evaluators it was important to take the concerns and arguments of the evaluatees seriously and try to include or at least consider them in the evaluation. On many occasions, the evaluatees were asked to provide relevant publications and other research material in order for the evaluators to become sufficiently informed. The evaluators were then given time to read the material to inform themselves about the evaluees and their products. Then, each researcher and research group was visited for presentations and informal talks. In due time, a preliminary formal report was drafted. An important feature is that the preliminary report was circulated to affected researchers whose written comments were explicitly solicited, and these comments were subsequently paid attention to when the evaluators composed the final report. However, they were not published along with the finished report. In the Scientific Commission of the National Swedish Building Council, the whole procedure took approximately eighteen months.

Collegial evaluation might also be pursued by use of two panels instead of one. These groups might work independently but in the end try to reconcile their findings and judgments. A special form of collegial evaluation is self–evaluation by the affected professionals them-

selves. Usually this approach is combined with the use of external evaluators.

Peer reviews frequently produce shaky results. Matched panels use widely different merit criteria and performance standards and reach miscellaneous conclusions. However, in technically complex fields, collegial evaluation is probably the finest method available to judge the quality of what is produced.

## Final Note

My broad survey of evaluation models has demonstrated that the total agreement that once existed in the early American and European evaluation communities on the appropriateness of the goal–attainment model has been replaced by a situation where several models compete. Internationally, evaluation has evolved from uniformity to pluralism. I greet this development with satisfaction. The danger with all evaluation models is that they are applied too uncritically and that decision–makers wrongly believe that one model can provide comprehensive, final answers. Therefore, it is important to keep in mind they every model provides partial perspectives and answers only. For this reason, combinations of several models is commended.

There is a strong tendency in contemporary evaluation literature to recommend stakeholder and client–oriented evaluation and debase particularly the goal–attainment model. On one important account, I take exception to the tendency of debunking the goal–attainment model. From a democratic point of view—and here I am alluding to representative democracy—the goal–attainment model and particularly the side–effects model are very important, since they are based on the conception of the parliamentary chain of influence. High–level policy goals, set by parliaments and governments are not just any goals whatever. Established through a constitutionally determined procedure, they are institutionalized as the collective goals of the state. Citizens and elected officials have legitimate reasons to ascertain whether policy goals have in fact materialized in the field. Otherwise, they cannot function as principals in the representative system of government.

The democratic argument in favor of the goal–achievement model, however, cannot remedy the fact, that this model runs into difficulties particularly with goal catalogues and is blind to costs. Its major drawback, however, is its lack of focus on side effects. For this reason, I prefer side–effects evaluation to goal–achievement evaluation.



Client–centered evaluation has a role to play, particularly as far as government services are concerned. However, for democratic reasons it cannot replace approaches that take policy goals as their organizer. The strength of peer review lies in its capacity to capture and judge qualities. This is essential in fields dominated by complicated criteria of merit and strong professions. Its paradigm area of application is academic research. The stakeholder model provides the broadest view possible of government interventions, promises to take all involved into consideration, but is impractical to handle. The Swedish version of the stakeholder model entails an interesting combination of social research and political accommodation of various stakeholder interests in order to shape social agreement and render legitimacy to decisions.

Economic models will stay with us forever in public policymaking. It must be kept in mind, however, that like other designs they provide partial perspectives only. The danger with economic models is that decision–makers are fascinated by their mathematical precision and wrongly believe that they provide comprehensive, final answers.

## Notes

1. Accounts of evaluation models are provided in Madaus et al. 1983, Guba and Lincoln 1981, and House 1980. In Shadish Jr. et al. 1991, evaluation models developed by Weiss, Wholey, Scriven, Rossi, Cronbach, and Stake are analyzed.
2. Goal–achievement: Scriven 1991:178; goal–based: see House 1980:26ff; goal-attainment model: Kaufman and Thomas 1980:126f.; objectives–oriented: see Guba and Lincoln 1981:x; the Tylerian model: Guba and Lincoln 1981:3ff., House 1980:27; the behavioral objectives approach: House 1980:26ff.
3. For criticism of the goal–attainment model, see Deutscher 1976, and Meyers 1981:110ff. The exposition in Chen 1990:168ff. is very clear and interesting.
4. By focusing on subject–matter goal achievement only, the goal–attainment model pays no attention to procedural goals. Procedural goals include legality and equity of client treatment.
5. Some people use the terms "internal and external effects, where the former accrue directly to the project, for example, elimination of mosquitoes as an internal effect of a mosquito–control project, with opening an area for recreation as an external effect" (Anderson and Ball 1978:26).
6. In Boudon 1982 (5 ff.) the expression "perverse effects" is used, but in the sense of unintended effects, which of course includes side effects as well.
7. Strategic effects cause some problems in my discourse on side effects. The desired main effect with the program could be that the party in government wins the next election. A possible side effect could be that the opposition splits up. In this section, however, I shall only consider substantive effects.
8. Ferguson's memorable phrase "results of human action but not of human design"

seems to have been phrased in the following fashion: "Every step and every movement of the multitude, even in what are termed enlightened ages, are made with equal blindness to the future; and nations stumble upon establishments, which are indeed the result of human action, but not the execution of any human design. If Cromwell said, That a man never mounts higher than when he knows not wither he is going; it may with more reason be affirmed of communities, that they admit of the greatest revolutions where no change is intended, and that the most refined politicians do not always know wither they are leading the state by their projects" (quoted from Hayek 1978:264).

9. Albert Hirschman (1991:35f.) has formulated the same idea in the following fashion: "One of the great insights of the science of society—found already in Vico and Mandeville and elaborated during the Scottish Enlightenment—is the observation that, because of imperfect foresight, human actions are apt to have unintended consequences of considerable scope. Reconnaissance and systematic description of such unintended consequences have ever since been a major assignment, if not the raison d'être, of social science." Also consult Boudon 1982; Vernon 1979; Sieber 1981; Hayek 1979:146ff; Elster 1978:106ff.
10. A neologism in the English language, *evaluand* is a generic term for whatever is evaluated, by analogy with "analysand" and "multiplicand." In this book, I have used "evaluand" only sparingly.
11. Other elaborated models of comprehensive evaluation include the ones elaborated by Fernández–Ballesteros 1992a:205 ff. (figure on 207) and Rossi and Freeman 1989:13f. and passim.
12. The presentation of the SOU–model draws on Söderlind and Petersson 1988; Meijer 1956; SOU 1976:49; Premfors 1983; Andersson and Associates 1978:55ff.; and Johansson 1992. For a short presentation in English, see Vedung 1992:76f.

# 5

# The Eight Problems Approach to Evaluation

Evaluation is a technique of managing public organizations in which systematic data gathering and other researchlike operations per definition play a considerable role. From this angle, research–infused evaluation differs from other social research not in regard to the research designs and data–collection methods employed, but to the problems to be attacked. Problems, not designs or methods, provide identity to evaluation.

There are eight primary problems of evaluation; they can be phrased as eight questions. Again inspired by Scriven, I shall call this the *Eight Problems Approach to Public Policy Evaluation.*[1] While question one and two concern the evaluation of the intervention, questions three through seven pertain to the intervention proper, and question eight to the feedback process or the utilization aspect of the evaluation. This three–level idea in the battery of questions, illustrated already in figure 2.2, ought to be kept in mind when we proceed.

Evaluators may consider some or all of the following eight problems:

1. The purpose problem: For what overall aims is the evaluation launched?
2. The organization (evaluator) problem: Who should exercise the evaluation and how should it be organized?
3. The intervention analysis problem: How is the evaluand, that is, the government intervention, normally the policy, the program, the components of policies and programs, or the provision of services and goods,

seems to have been phrased in the following fashion: "Every step and every movement of the multitude, even in what are termed enlightened ages, are made with equal blindness to the future; and nations stumble upon establishments, which are indeed the result of human action, but not the execution of any human design. If Cromwell said, That a man never mounts higher than when he knows not wither he is going; it may with more reason be affirmed of communities, that they admit of the greatest revolutions where no change is intended, and that the most refined politicians do not always know wither they are leading the state by their projects" (quoted from Hayek 1978:264).

9. Albert Hirschman (1991:35f.) has formulated the same idea in the following fashion: "One of the great insights of the science of society—found already in Vico and Mandeville and elaborated during the Scottish Enlightenment—is the observation that, because of imperfect foresight, human actions are apt to have unintended consequences of considerable scope. Reconnaissance and systematic description of such unintended consequences have ever since been a major assignment, if not the raison d'être, of social science." Also consult Boudon 1982; Vernon 1979; Sieber 1981; Hayek 1979:146ff; Elster 1978:106ff.
10. A neologism in the English language, *evaluand* is a generic term for whatever is evaluated, by analogy with "analysand" and "multiplicand." In this book, I have used "evaluand" only sparingly.
11. Other elaborated models of comprehensive evaluation include the ones elaborated by Fernández–Ballesteros 1992a:205 ff. (figure on 207) and Rossi and Freeman 1989:13f. and passim.
12. The presentation of the SOU–model draws on Söderlind and Petersson 1988; Meijer 1956; SOU 1976:49; Premfors 1983; Andersson and Associates 1978:55ff.; and Johansson 1992. For a short presentation in English, see Vedung 1992:76f.

# 5

# The Eight Problems Approach to Evaluation

Evaluation is a technique of managing public organizations in which systematic data gathering and other researchlike operations per definition play a considerable role. From this angle, research–infused evaluation differs from other social research not in regard to the research designs and data–collection methods employed, but to the problems to be attacked. Problems, not designs or methods, provide identity to evaluation.

There are eight primary problems of evaluation; they can be phrased as eight questions. Again inspired by Scriven, I shall call this the *Eight Problems Approach to Public Policy Evaluation.*[1] While question one and two concern the evaluation of the intervention, questions three through seven pertain to the intervention proper, and question eight to the feedback process or the utilization aspect of the evaluation. This three–level idea in the battery of questions, illustrated already in figure 2.2, ought to be kept in mind when we proceed.

Evaluators may consider some or all of the following eight problems:

1. The purpose problem: For what overall aims is the evaluation launched?
2. The organization (evaluator) problem: Who should exercise the evaluation and how should it be organized?
3. The intervention analysis problem: How is the evaluand, that is, the government intervention, normally the policy, the program, the components of policies and programs, or the provision of services and goods,

to be characterized and described? Is the evaluand regarded as a means or as a self–contained entity?

4. The conversion problem: What does execution look like between the formal instigation of the intervention and the final outputs?
5. The results problem: What are the outputs and the outcomes—immediate, intermediate, and ultimate—of the intervention?
6. The impact problem: What contingencies (causal factors, operating causal forces)—the intervention included—explain the results?
7. The criterion problem: By what value criteria should the merits of the intervention be assessed? By what standards of performance on the value criteria can success or failure or satisfactory performance be judged? And what are the actual merits of the intervention?
8. The utilization problem: How is the evaluation to be utilized? How is it actually used?

Questions one and two concern the evaluation of the intervention, three through seven the intervention itself, and question eight the feedback process.

Since evaluation is frequently made to order, it is advisable to consider the overall *purposes* of the commissioner. True, the evaluator must find out what policy, program, or program ingredient the sponsor wants to have investigated. Furthermore, the evaluator should also pay attention to the particular problems that the commissioner wants her to illuminate. In talking about evaluation purposes, however, I have something more far–fetched in mind, namely, why the sponsor wants to have these particular things examined. In which decision context will the evaluation be used? Who are the prospective primary users? Is there a hidden agenda behind the evaluation? Is there really a genuine desire with the sponsor to have the program and its outputs and outcomes clarified or does he covet a tranquilizer or some legitimizing evidence?

The purpose problem in the overall sense intimated here is something the basic researcher may circumvent. In the academic community, funds for research are allocated on the grounds of basic–knowledge interests alone. Not so in evaluation. A client commissioning an evaluation also has a political agenda in mind. Hence, the evaluator must pay attention to the commissioner's deeper purposes.

As a second step, it is appropriate to ponder the issue of how evaluation work should be *organized*. Who should conduct the evaluation? The organization problem can be viewed from different angles, which I have touched upon in chapter 4 in conjunction with my discussion of the Swedish policy–commission model and the peer review



model. Here, I shall briefly address only the issue of self–evaluation versus external evaluation. Should the evaluation be initiated and produced by the affected people themselves or should some external body commission and conduct the assessment?

At an early stage, the evaluator must make herself familiar with the *intervention*. How should the intervention—policy, reform, plan, program, program component, services, products—be described? What is the nature of the intervention? How should it be depicted? Many interventions are exceedingly diverse, varying substantially from site to site throughout the country.

Occasionally, the issue of describing the intervention can be lightly dealt with. This is true, of course, in self–evaluation conducted by people who already know the program well. Also in externally performed client–centered and stakeholder–oriented evaluation, it is perhaps less essential to dwell upon the nature of the program. In side–effects evaluation, which I prefer to goal–attainment and comprehensive evaluation, the program issue gains in importance. Here, it may be proper to analyze the intervention in terms of ends and means. What goals are laid down in the policy mandate? If several goals are set, how are they ranked? What range of policy instruments are incorporated in the program? Does the program entail regulatory, economic or communication tools of governance? In case several instruments are involved, how are they combined? If only one type of policy instrument is devised —for example, regulations—what kinds of regulations are they?

At this juncture, I must draw a contrast between two cardinal types: explanatory and nonexplanatory evaluations. *Explanatory evaluation* amounts to means–ends evaluation. It sets out to assess programs with respect to something they are supposed to produce or have inadvertently generated. It raises the conversion issue and maybe the impact issue as well, that is, the evaluator attempts to mirror the conversion process, and encircle the outputs and the outcomes of the program; the focus of attention is not put directly and only on the program, but on its implementation and results. *Nonexplanatory evaluation,* on the other hand, assesses the merit, worth, and value of the program through direct appraisal of the program itself, not through assessment of something produced by it. There is no causality issue involved in nonexplanatory evaluation.

For instance, a students' appraisal of a university course is a