

## INTEGRATED EVALUATION: A SYNTHESIS OF APPROACHES TO THE EVALUATION OF BROAD-AIM SOCIAL PROGRAMS

RACHELLE ALTERMAN, NAOMI CARMON and MOSHE HILL

Faculty of Architecture and Town Planning and Samuel Neaman Institute for Advanced Studies in  
Science and Technology, Technion—Israel Institute of Technology, Haifa 32000, Israel

**Abstract**—Integrated evaluation is intended to serve decision makers who are responsible for broad-aim social programs by providing information based on evaluation which can aid both ongoing decisions and long-term strategic decisions. It integrates elements of diverse evaluation traditions in a complementary manner: monitoring—in order to inform what has been done by the program; implementation analysis—in order to understand how decisions are being made and carried out; economic evaluation—including both cost-effectiveness and assessment of distributional effects; and goal achievement evaluation—in order to present the program outcomes from the point of view of the various parties who produced the program and/or were affected by it. The article presents these four components of integrated evaluation and discusses its advantages as well as its difficulties and pitfalls.

### INTRODUCTION

Evaluation research has been on the scene for several decades. During this period, it has found a home in the behavioral sciences, has developed an impressive methodology, has been commissioned by thousands of decision makers and applied to a wide range of social programs. However, for the past few years, evaluation research has been faced with a growing wave of criticism. It has been attacked for being oblivious to the needs of the decision makers, often remaining unused; for taking too long and costing too much [1]; for ignoring the goals of participants other than high level officials, especially the goals of program recipients [2]; for assuming a set of fixed goals which in practice soon drift along and become remolded [3]; for insisting on experimental or quasi-experimental rigorous designs which often prevent the assessment of what really happened in the field; and for neglecting the use of social theory [4].

Recently, one can identify the contours of what has been called the emerging “revisionist” approach [5]. It is more “utilization focused” [6]; it emphasizes process, and not only outcome [7]; it does not rely on fixed goals [8]; and it is more qualitative than quantitative [9, 10].

This emerging approach to evaluation research bears obvious kinship to the rapidly developing field of implementation analysis which seeks to study the factors affecting the likelihood of successful implementation [11]. Yet the two fields have developed in separate tracks, the former usually being based in sociology and psychology and the latter having its base in policy studies and political science. Only recently have the two trends begun to show some mutual recognition [5, 12].

In addition, both the well-established ex-post approach of the behavioral scientist and the *process* approach of the political scientist usually ignore two other important traditions: that of *ex-ante* evaluation which has been developed by economists and urban

planners [13-16], and that of *monitoring*, commonly applied by managers. These two have as yet not taken direct part in the current debate.

The growing attack on the mainstream approach presents the danger of a divergence between it and the contending views, whose outcome may well be a growing reluctance among decision makers to continue the funding of evaluation research. The purpose of this article is to attempt to prevent this scenario from happening. By proposing the concept of “integrated evaluation”, this study seeks to demonstrate how the various approaches to evaluation may complement each other in attempting to answer those questions about social programs that interest decision makers. It is felt that only by drawing upon the rich set of traditions in evaluation research, and by integrating them properly so as to meet the well-justified criticisms mentioned previously, will the task of evaluating social programs be advanced.

The approach proposed is tailored particularly to the evaluation of broad-aim social programs. By this term we mean to include programs such as a comprehensive program of neighborhood rehabilitation or the introduction of major changes in a health system. Broad-aim programs are often instituted on a national or regional level and may be viewed as “macro” rather than “micro” programs in that within each one there may be many particular programs with distinct sets of goals. The planning and implementation processes of such programs are likely to be dynamic, with the continuing involvement of numerous actors. These actors strive to influence the goals of the program and its detailed implementation. Moreover, the process is likely to be iterative. The interests of these groups are not necessarily fixed and their goals are likely to change in the course of time, particularly after the implementation has commenced and its effects are being felt. Hence, if the evaluation study were to take into account only the intended goals of the program initiators, it would at best be irrelevant and at worst misleading.

### THE CONCEPT OF INTEGRATED EVALUATION

The proposed approach is integrated in two senses. First, it integrates elements of various evaluation traditions, relying on both the traditional *ex-post* approach (with some important modifications) and some of the newer approaches, as follows:

- It utilizes the goals achievement matrix<sup>†</sup> applied mostly for *ex-ante* evaluation<sup>‡</sup> but useful for *ex-post* evaluation as well.<sup>§</sup> That is, it is concerned with the effects of the program on the multiple interest groups affected by or influencing the program. It is thus particularly oriented to the distributional effects of the program, enquiring also after the relative importance of the goals from the points of view of the groups involved.
- It uses monitoring, an approach developed in the field of management, in order to measure what is being done on an ongoing basis.
- It incorporates the concepts being developed recently in "implementation analysis" or "implementation process evaluation", in order to investigate how decisions are being carried out.
- It records economic costs and relates them to outputs and outcomes for purposes of cost-effectiveness analysis.
- It uses the measurement techniques which have been developed by behavioral scientists and widely used in social impact assessment studies, and it also uses social theory at several critical points in the evaluation study.

Secondly, it is also integrated in the senses of fitting into the decision-making process on an ongoing basis rather than on a one-shot basis as the traditional approach does. It accomplishes this in the following ways:

- It accompanies the continuing process by providing answers which can aid both ongoing and periodic decisions.
- It is attuned to the dynamic and iterative nature of the planning and implementation process of broad-aim social programs.
- It involves the decision makers in the evaluation process and provides them with the results in a manner to which they can easily relate.

<sup>†</sup> The goal achievement matrix was developed by Hill [17]. See also Hill [13].

<sup>‡</sup> This approach has often been employed by urban planners [18-20].

<sup>§</sup> Rossi *et al.* [21] are among the few analysts who view the various approaches not as alternatives, but rather as complementary to each other. Each approach is seen as answering different questions that arise at the various phases of the planning and implementation process of deliberate social change. What they term "comprehensive evaluation" is based on the sequential application of types of evaluation, whereby the need for subsequent types of evaluation is justified by the findings in the previous types of evaluation. For example: they argue that there is no reason to conduct impact assessment unless the monitoring of the implementation indicates that the outputs in the field were indeed in accordance with the plan. This approach ignores the fact that sometimes the mere declaration of intent to implement a program may lead to significant costs and outcomes, even if it was not implemented, or if only a small part of it was in fact implemented.

### THE QUESTIONS OF INTEGRATED EVALUATION

The proposed evaluation approach is intended to provide answers to the following questions:

- (a) Is the program being implemented? To what extent do the ongoing outputs and costs of the interventions in the system comply with the guidelines of the plan?
- (b) What are the economic costs? Who are the groups who are bearing the costs of the program? Does the actual distribution of the costs accord with that specified in the initial plan?
- (c) Is the implementation process effective? or: to what extent do the political and administrative structures and decisions in the course of the implementation of the program advance the prospects that the implementation will be in accordance with the planned intervention and its goals?
- (d) What are the program outcomes and who benefits from them? or: to what extent do the outcomes of the program contribute to the achievement of the important goals of each of the publics and interest groups who influence the program or are influenced by it? Who are the main beneficiaries of the program? Is this in accordance with the plan?
- (e) What are the social, economic, political, and administrative conditions which have enhanced or have prevented the achievement of the program goals?

The answers to the first three questions are essential for ongoing decision-making in the course of the implementation of the program; in practice they can provide an "early warning system" that is likely to signal possible deviations from the plan and its goals at an early stage.

The answers to all these questions and particularly to the two latter ones are intended to provide the basis for periodic decisions with respect to the continuation of or the cessation of the program, the introduction of changes into it and/or its expansion in order to serve additional population groups.

### THE COMPONENTS OF INTEGRATED EVALUATION

Integrated evaluation is composed of four components. We prefer to use the term "components" rather than "stages" because we believe that the decision-making process is seldom sequential and thus a strict order in the evaluation process should be avoided.

The four components of integrated evaluation are:

- (1) Monitoring of outputs and costs.
- (2) Implementation process evaluation.
- (3) Economic evaluation: cost-effectiveness analysis and evaluation of distributional effects.
- (4) Evaluating program outcomes from a multi-group perspective.

A description of each component now follows.

#### *Monitoring of outputs and costs*

Monitoring is the ongoing feeding of information to the decision-making process about outputs in the field, i.e. services delivered. Although the term is

variously used,† in this article we shall use it as it pertains to the gathering of information about costs and outputs‡ only, so as to distinguish it from implementation process evaluation, economic evaluation, and evaluation of outcomes.

Systematic monitoring is necessary in order to check the extent to which the program delivery is in accordance with the program plan in terms of the substantive outputs, their location, their timing, and their costs. This information is necessary for the ongoing management of the program in the field, because it helps to keep the implementation in line with the program specifications. It may also aid those responsible for making periodic decisions about the continuation of the program.

The questions to be answered by systematic monitoring are:

- Who should have carried out what, when, where, and for whom? Or: what has been planned?
- What is being delivered? Or: what are the program outputs?
  - To whom is it being delivered? To which types of people, organizations, or sites is it being delivered?
  - When is it being delivered? Or: how long does it take to start the delivery and how long does it take to deliver?
  - How much does it cost in monetary terms?
  - Who pays the costs? Which of the groups and publics who influence the program or are influenced by it pay how much of the costs?
  - How do the above findings compare with the planned outputs and costs?

Monitoring of outputs and costs can be designed in many different ways,§ differing in breadth, scope, and costs in accordance with the objectives and resources of a particular evaluation study. It can be designed as part of a “rapid feedback evaluation”||

---

† Various uses of the term in urban planning are reviewed by Alterman [11]. Often one reads references to “implementation monitoring” which is sometimes used to mean monitoring of outputs [22], and is sometimes used to refer in addition to some of the aspects of implementation process evaluation described below [21].

Another term used in Britain is “impact monitoring” which deals with questions of outcomes, similar to those dealt with by ex-post impact assessment, but on an ongoing basis. A third term is “strategic monitoring” which draws on ‘feedforward’ information for the planned system and is supposed to guide changes in policy. And finally, Wholey [2] talks of “performance monitoring” which he defines as “the periodic measurement of progress toward program objectives.” This definition does not help us distinguish between the monitoring of outputs and of impacts (or ‘outcomes’) and it is thus difficult to separate it from impact evaluation.

‡ Outputs differ from outcomes: outputs are expressions of performance of the delivery system (e.g. number of teaching hours), while outcomes are the consequences of the performance (e.g. students’ achievements). As Weiss and Rein [23] have argued effectively, politicians and decision makers are often more concerned with outputs than with outcomes: the latter are sometimes too elusive, invisible, and open to interpretation, while the former are concrete, visible, and factual.

§ Based on Rossi *et al.* [21], Chap. 4, and on Wholey [2] Chap. 12.

|| As termed by Wholey [2].

using suitable low-cost data sources such as: site visits; telephone surveys; interviews with key delivery system personnel and opponents of the program; perusal of newspaper items and other publications; and data from audits and budgets. Or it can be designed as a systematic information system, feeding in more reliable and comprehensive information, based on sources such as questionnaires or interviews with service personnel; collection and analysis of administrative and service records (reports filled routinely by service personnel about services rendered, when, to whom etc.); survey of a sample of target groups regarding extent of participation in the program; systematic observational data on outputs provided by observers stationed at strategic locations; and systematic data on expenditures.

#### *Implementation process evaluation*

Implementation process evaluation focuses on the political and administrative processes occurring within the program delivery system and their relationships with the program plan and the program goals.† It asks these major questions:

- How does the implementation take place? Who are the political and bureaucratic parties that are involved in the implementation process? How do they interact and what decisions do they make?
- How effective is the implementation process? In other words: do the decisions being made enhance the chances that in the final analysis the program will be performed in accordance with its plan and with its goals?

So-called process evaluation [23], as usually practiced, asks only the first question. However, answering it tends to yield an avalanche of descriptive data of the “what happened” and “who said what to whom” genre. Therefore, we suggest restricting the collection of descriptive data to those which are necessary for answering the second question. Such data may indicate which characteristics of the various groups participating in the implementation process affect the likelihood of compliance of subsequent decisions on guidelines or allocations with the initial program goals. Here one might take into consideration such factors as the relative degree of power of the various groups, the degree of commitment of the various types of personnel, effectiveness of coordinating mechanisms, types of controls and incentives, financial resources available, etc. [25].††

The rudimentary state of the art of implementation process analysis does not as yet provide us with a set

---

† We did not find in the literature on evaluation a precise parallel to our approach. However, various authors have employed one or another aspect of our approach and called it by different names. Weiss and Rein [23] talk of “process-oriented qualitative research”; Wholey [2] talks of “administrative monitoring”, and “performance monitoring”; Rossi, Freeman, and Wright [21] talk of “monitoring program implementation” in which they deal with some of the questions posed here, plus some posed under the section dealing with “monitoring of outputs”. The closest concepts to our own are stated by Perkins [24], when he discusses “compliance evaluation” and “management evaluation”.

†† See also Van Horn [26]; Mountjoy and O’Toole [27]; Edwards [23]; and Nakamura and Smallwood [29].

of distinct and tested methods whereby these questions can be answered. In the absence of generalizable knowledge about factors affecting implementation, the exact questions posed and the methods suitable for tackling them will have to be tailored to each case; in fact, this is as it should be, considering that each implementation setup and process is unique. However, we can provide several clues about useful approaches:

- A helpful initial approach is to draw up a comprehensive checklist of questions and factors to be studied such as those posed above, tailored to the particular attributes of the implementing institutions and processes in question. Often, initial answers provided to such a checklist by the analysts or through interviews with knowledgeable persons within and outside the implementing system can go a long way without extensive systematic empirical research, and can serve as part of a "rapid feedback" evaluation system.

- An approach which has gained some popularity is the "narrative" approach [23], or the "action-response" approach [30]. This is mostly a description of what is going on and who is doing what on a day-to-day basis, with some attempt to understand why the occurrences happen as they do.† According to our judgment, this framework makes it difficult to go beyond a description of the implementation process.

- An approach which can take us beyond descriptive analysis is to draw up a "logic model" of the implementation process [2, 31], using a theoretical framework which can be drawn from various studies of implementation processes: (1) the implementation process as a game (or set of games) among the various groups of "players", each attempting to maximize its own goals [32]; (2) the implementation process as a chain of decisions at "clearance points" to be made by various institutions and actors, each linked to the other with a certain probability of compliance, where the outcomes are the cumulated conditional probabilities [33]; and (3) the implementation process as a control system with subsystems and relationships among them, having a set of functions to be filled [23, 34]. All these various approaches could use the following types of data sources: internal documents and protocols, interviews with group representatives and service personnel. Wherever needed and possible, these approaches could use "participant observers" who report on the participants and the processes of decision-making.

In traditional evaluation research, the implementation process is a "black box". By opening up this box, much of the problem of determining causal linkages is by-passed. Instead of the distant program-to-outcome relationship, a set of closely-related linkages is exposed, as will be further explained later.

#### *Economic evaluation*

In practice, when traditional evaluation of outcomes is commissioned, economic evaluation is either not included at all or is commissioned separately. Decision

makers, however, usually find this type of evaluation crucial to public decision making. We have therefore included it as an essential element in our approach to evaluation and have emphasized the importance of linking it with monitoring and the evaluation of outcomes from the start.

The tools of the economist are employed to evaluate two central aspects of broad-aim social programs—the efficiency of resource allocation and distributional equity.

Cost-effectiveness analysis [35, 36] measures the efficiency of resource allocation in the delivery of outputs, i.e. services delivered and/or efficiency in achieving desired outcomes measured in terms of goal-achievement. The cost measure expressed the monetary value of the resources that have been employed, such as labor, equipment, and physical facilities. The effectiveness measure refers to units of output or units of outcome. For example, in the case of an educational program, typical output measures may be hours of instruction and numbers of beneficiaries or hours of instruction per beneficiary. Measurement of program outcomes might be in terms of performance of beneficiaries in scholastic achievement tests or in the extent of reduction in the school dropout rates of the beneficiaries.

Cost-effectiveness analysis may be employed for various purposes. The costs of similar levels of output at various locations where services have been delivered can be compared in order to establish relative efficiency. The costs of service provision at various levels and intensities of output can be compared in order to establish marginal levels of efficiency thereby indicating an optimal level of output from the efficiency point of view. If different program elements are intended to serve the same goal, cost-effectiveness analysis enables us to compare and rank each program in terms of its relative costs for the same level of goal achievement.

Program outcomes can be expressed in economic terms, if outcomes are assessed in market prices, or in terms of the amount of money that the beneficiaries would be willing to pay in order to achieve the desired outcomes. In this case both costs and effectiveness can be expressed in monetary terms (with measures of outcome now termed benefits) and it is possible to carry out a cost-benefit analysis. A measure of net benefit (benefit minus cost) expresses the net economic value of the program to society and is hence a more adequate measure of economic efficiency than is provided by cost-effectiveness analysis. However, for most social programs it is not possible to measure program outcomes in economic terms since outcomes are not measurable in market prices or subject to the willingness to pay criterion for assessment of benefits. In this case cost-benefit analysis is precluded [37].

Distributional equity is usually a central concern of broad-aim social programs. The analysis of equity focuses on the distribution of costs and benefits. It examines the questions of who bears the cost of the program and who benefits from it. A central concern is whether the costs and the services delivered by the program are distributed in accordance with its objectives. It may examine issues of target efficiency.‡

† This approach has been applied to some of the evaluations undertaken on the Model Cities Program by Marshall Kaplan, Gans and Kuhn.

‡ See Campbell and Stanley [38]; Rossi and Williams [39]; Hatry *et al.* [40]; and Riecken and Boruch [41].

such as the extent that the program is reaching those socio-economic groups to which it has been targeted. The question here is what proportion of the target population has been reached and which part is not benefitting from the program. A related question is which part of the services, delivered by the program, benefits sections of the population for whom the program was not intended. Another distributional question is: what is the equivalent increment to income that has been provided to the beneficiaries of the program through the services that have been delivered to them?

*Measuring program outcomes from a multi-group perspective*

Measuring program outcomes is often the focus of evaluation research. The usual way of doing this is by identifying the goals of the responsible authorities, expressing the goals operationally, and measuring goal achievement. The customary method is a quasi-experimental research design which provides a basis for tracing cause-effect relationships between the program under study and the outcome variables.

The approach that we have adopted differs from the traditional research methods in two important respects:

- The point of departure is not only the set of goals of the authorities but the goals of all the parties who are interested in the programs or are affected by it.
- We propose to use neither experimental nor quasi-experimental methods for causal analysis.

The identification of all the publics who are involved in the plan is likely to be a complex task. Among these publics are the governmental bodies who are the immediate sponsors of the plan and other agencies of central and local government, groups who are the immediate beneficiaries of the program and other publics who are interested in the program or who are influenced by it. In every case where the situation so enables, representatives of the most relevant publics or interest groups should be consulted by the analysts in order to identify their most important goals, by order of preference.

The manner in which goals are articulated by the various groups will undoubtedly differ from case to case. In some cases, the goals will be expressed in measurable terms; in other cases the goals expressed will be more general. Their reduction into operational objectives will have to be undertaken by the analysts based on their understanding of the validity of particular indicators as expressing the essence of the goals, and the reliability of the measures under various conditions. Whenever possible the goals will be expressed quantitatively, e.g. housing density, number of high school graduates, etc. Failing this, the achievement of qualitative goals can be measured ordinally in terms of judgmental criteria, e.g. goals relating to visual quality, or neighborhood image. The operational definitions of the important goals of the major groups who are interested in the programs or affected by it constitute the set which will be measured by the evaluation study.

In order for the measured changes to be considered as the program outcomes (and not as consequences of other events which took place at this same time) one must ensure three conditions [12]: the correct

chronological sequence of the appearance of the variables (the variable identified as the cause must appear before the one identified as the effect); correlated change (when one variable changes, so must the other); and elimination of other factors that may have caused the same effect. It is relatively simple to ensure the existence of these three conditions in an experiment. However, by contrast with many other analysts, [46] we do not recommend experimental or quasi-experimental study of carefully selected experimental and control groups as a way of dealing with this dilemma. † An important reason for this is related to the fact that in many broad-aim social programs the appropriate unit of analysis is large: not an individual or a household, but rather an urban neighborhood, a minority group in the center of the city, etc. When the unit of analysis is so large, it is almost impossible to locate sufficient similar units to enable inferences about a large number of relevant variables. Moreover, since complex implementation processes can seldom be replicated, there is no way of ensuring similar implementation of broad-aim programs in the different units of analysis.

Therefore, we suggest the method of 'before' and 'after' measurement with continuous observation of the resource inputs and program outputs and the implementation processes relating to these for the (large) units of study. The 'before' and 'after' measures ensure the ability to determine chronological order and to identify correlated change. The continuous observation and analysis, as expressed in the monitoring and implementation process evaluation, opens up the 'black box' relating inputs to outcomes and reveals the connections between them in what Thomas [7] calls a 'close causation' approach.

As far as possible, the tools of measurement should be simple and inexpensive. Sometimes there may be no escape from repeated household surveys, but wherever possible it is desirable to restrict this expensive means of measuring outcomes. Instead, it is recommended that available data should first be scanned. Usually, much data are available from the agencies connected with the program. These sources could be utilized in their existing form, or they could be expanded to include information necessary for purposes of evaluation. The involvement of the program personnel in the evaluation process can lead to gaining their support and so increasing the likelihood that they will be willing to utilize the results of the evaluation.

Once the outcomes are measured, they are related to the goals of the relevant publics. Thus the analyst should be able to connect the various threads and to provide the decision makers with a goals-achievement matrix (Exhibit 1). This matrix provides a summary statement of the contribution of the program under consideration to the achievement of the goals of the various publics and interest groups involved in the program.

---

† The idea of comparing an experimental unit with a control unit (in medicine, for example) mainly resulted from the fact that it was only possible to investigate the inputs and outputs and not the process leading from one to the other. Since in our case it is proposed to study the implementation process in each case, the investigation of the outcomes in control units is less essential.

**INTEGRATED EVALUATION—DIFFICULTIES AND PITFALLS**

As with all methodologies, integrated evaluation has its characteristic difficulties and pitfalls, only some of which can be satisfactorily countered. We shall first review the difficulties for each of the components and then for the methodology as a whole.

A major problem in monitoring costs and outputs is the tracing of budgetary displacement. This frequently occurs in the case of broad-aim social programs in which many agencies are involved. An agency may take a ride on a social program provided by another agency which is intended to be a net addition to an existing program. Budgets provided for the additional new program may be siphoned off to the existing program. Such displacement is difficult to trace since the agencies try to disguise it because it is contrary to administrative directives. In cases like this it may be difficult to point at the real economic costs of the program. However, when this occurs it may not be a net addition to the costs of the public fisc. If a project that was previously funded from other sources is now funded by the new program, the funds from the previous source are now available for other public purposes. In this case there is administrative malpractice but there are no additional economic costs.

On the output side, i.e. service delivery, for broad-aim programs, one is faced with great breadth of data to be collected from many agencies. The effort and cost of assembling this may be considerable. If one relies only on data collected by the agencies themselves, there may be problems of reliability in the information due to the 'positive' bias (tendency to exaggerate) of those responsible for service delivery when reporting to higher authorities. The consumers of services, on the other hand, may have difficulty in differentiating and attributing the sources of services that they receive.

Implementation analysis is a methodology in the making and the state of the art is still formative and has not been widely tested. There are certainly difficulties that have to be faced in its application. It is essential to have access to inside information and, for this, cooperation of the authorities is necessary. Many agencies are probably involved or affected by broad-aim programs and it is quite unlikely that all of them will be supportive of the evaluation.

Another inherent difficulty is that, inevitably, all the informants have their particular interests, are not

objective, and it is therefore difficult to arrive at an accurate and agreed-upon assessment of factors likely to affect success or failure. The broader the program the more actors involved and the more complex the webs of connections and influences affecting courses of action. Implementation analysis is based on qualitative rather than quantitative information. Different people may consequently make different judgments on the basis of the same qualitative information. Another potential pitfall is the difficulty of separating out the effect of individual personalities from the effects of other, more structural, factors affecting the implementation process.

Cost-effectiveness analysis can only be as good as is enabled by the quality of both the monitoring of costs and outputs and the measurements of outcomes. The costs which are usually taken into consideration are the direct cost outlays. There may, however, be secondary costs which are not immediately evident. For instance, the provision of services to beneficiaries in a given program may cut down the demand for other existing related programs, resulting in inefficiency from the loss of scale economies and hence an increase in the relative costs of provision of their services.

The tracing of distribution effects requires information about target populations which may not always be readily available. Especially difficult to obtain is information about that part of the target population which is not benefitting from the program, and those beneficiaries of the program who are not part of the target population. A critical issue which has to be resolved is what is the ultimate criterion according to which distributional equity is assessed. Unlike the efficiency criterion which can be measured objectively, the very criterion for assessing distributional equity is subjective. Various equity criteria have been proposed [43, 44] such as need, equality, right, desert, etc., but the criterion of choice is ultimately the one that has been subjectively preferred by the decision makers. Consequently, the results of evaluation of distributional equity will vary according to the equity criterion that has been employed.

There are several potential problems with the fourth component of the evaluation—the measurement of program outcomes and their evaluation.

When one considers broad-aim social programs, the number of groups affected, whether directly or indirectly, is likely to be very large and it may be difficult to identify all of them. Moreover, it is reasonable to assume that often such programs affect

Exhibit 1. Goals-achievement matrix

	Goal A		Goal B		Costs as borne by groups
	Relative weight	Level of achievement	Relative weight	Level of achievement	
Publics and interest groups					
Central decision makers					
Local decision makers					
Service delivery personnel					
Client Group A					
Client Group B					
Other affected publics					

people who are difficult to get at, such as future residents of a neighborhood. Since it is not feasible to relate to all of them, the analysts have to rely on their best judgements and relate to the most relevant publics and interest groups, those for whom the implementation or the non-implementation of the program is particularly significant.

Another difficulty arises in identifying the goals of the various publics. It is not always obvious who speaks for the various groups and on what basis. The authorities who have initiated the evaluation might oppose the inclusion of the goals of certain publics which may be in conflict with those of authorized decision-making bodies. The identification of the preferences of the publics can also be problematic. Useful techniques for this purpose can be trade-off games, nominal group methods, etc.

We have suggested that various quasi-experimental methods are not suitable for evaluating the impact of broad-aim social programs on large research units (such as a neighborhood) and have proposed an alternative approach. It is clear to us, however, that the 'softer' methods of analysis that we have suggested might raise doubts about whether we can really identify cause and effect, at least in some of the cases. Moreover, while methods of continuous monitoring and the 'close causation' approach suggested by us are quite good for locating errors of Type 1 (an assessment that the program had an effect when it actually had none), they are less sensitive to errors of Type 2 (an assessment that the program did not have an effect when it actually did).

These are difficulties that arise from our attempt to assess the effect of programs on the multiple objectives held by multiple parties. There may be difficulty in disaggregating outcomes in order to determine their differential effects on different groups. On the other hand, when there are multiple effects, there is an inherent difficulty in arriving at an aggregate outcome measure for outcomes that are not commensurate. There is also a question of the trading off of outcomes measured on different scales and affecting different groups. Our approach is to avoid the aggregation of non-commensurate data by recording the disaggregated effects on all the parties in terms of the entire set of outcomes.

Turning now to general problems relating to integrated evaluation as a whole, we should first note that in order to implement the approach it is necessary to have suitably trained personnel, and considerable financial resources may be required. Therefore this approach is not particularly suitable for the evaluation of programs which are limited in their scope. It is primarily intended for broad-aim programs delivering services to many consumers. Almost every government introduces such programs, focusing on urban areas or on broad policy areas such as health policy, education policy, or social welfare. The total number of such programs is small but decisions made about them are of particular importance because they make large demands on public resources. The program may also involve a significant change in public policy, therefore investment in integrated evaluation of such programs would seem to be eminently worthwhile.

Even when resources for evaluation are available, other difficulties have to be overcome when carrying out this evaluation procedure.

First, integrated evaluation consciously assumes a trade-off between in-depth evaluative research in favor of in-breadth evaluative research. The broad scope that we have proposed comes inevitably at the expense of in-depth research on specific aspects. Furthermore the multidisciplinary character of the research team required to carry out integrated evaluation is likely to generate less in-depth analysis than research based on a single disciplinary point of view.

Integrated evaluation of broad-aim programs is likely to require at least two to three years to be completed. Frequently, quicker feedback is required from the evaluators. A partial answer to this problem may come through the provision of intermediate results based on the continuous monitoring of costs and outputs and the evaluation of the implementation process some time before the measurement and evaluation of outcomes have been completed.

Finally, this type of evaluation almost never arrives at a conclusion which states unequivocally whether the program was a success or a failure. Instead, the characteristic conclusion is likely to be that some of the goals of some of the publics have benefitted while the goals of other publics have been affected adversely or not at all. It may be more difficult for decision makers to wrestle with this type of conclusion even though it may be more comprehensive and balanced.

#### THE ADVANTAGES OF INTEGRATED EVALUATION

The integrated evaluation approach has several advantages over traditional approaches and especially over impact assessment [21], which is the most common evaluation procedure for health, education, and social welfare programs. Among its advantages are:

- It takes into consideration not only the aims of the program suppliers and producers, but also those of its consumers; not only the goals of the decision makers in the central government but also of local decision makers and of groups who do not have institutionalized political power.

- The approach takes into account the distributive effects of the program. It traces who benefits from the program and who bears the costs of the program. Not only does this express the equity of the program but it can assist in determining the political acceptability of the program and the likelihood of its continued implementation.

- It does not make the erroneous assumptions that programs have a single set of goals which are stable through time, and that these goals can be translated into agreed-upon measurable criteria which constitute a fixed set of appropriate reference points against which impacts can and should be evaluated. Instead, it recognizes the existence of several sets of goals which are likely to change in the course of the dynamic and iterative process of the planning and implementation of broad-aim social programs.

- It considers economic costs, but like cost-effectiveness analysis, and unlike cost-benefit analysis, it does not necessarily translate the effects into monetary terms.

- At least some of its results can serve decision makers in the course of the implementation process, creating an "early warning system". This can moderate

the contention that one must wait too long until the study results become available.

- The almost impossible task (in the case of large evaluation study units) of finding a sufficient number of matched experimental and control groups is avoided. Instead we open the "black box" of the implementation process and attempt to trace the connections between the inputs and the outcomes by continuous monitoring.

- The monitoring of the outputs and the evaluation of the implementation process significantly broaden our understanding of the reasons for success or failure of programs and therefore of the lessons that may be learned from them.

- By identifying the relevant publics involved and their goals, we create rich and wide-ranging sources of criteria for measurement of the success of the program; in this way the danger is reduced that the evaluation study will arrive at the erroneous conclusion that there has been no effect, a conclusion which might be reached if only the small number of goals of the suppliers are used as reference points against which outcomes are judged.

- The suggested criteria for measuring success is not restricted to the intended goals of the decision makers; if the analysts apply theoretical and empirical knowledge concerning the structure of the system and its behavior at the time of the identification of the relevant groups and their goals, the list of criteria derived from these goals will include what is usually known as "side effects" and "second-order consequences". By including the possibility that the list of groups and goals can be changed and modified in the course of the implementation, one assures that the most significant and important effects of the program can be accounted for.†

- Last but not least, integrated evaluation calls for the application of social science theory for the purpose of the identification of the relevant groups affected, for the formulation of their goals in operational terms, and especially for the specification of the goals of those affected who have no representation in the evaluation process. Hence, integrated evaluation is not only an application of social research techniques (as Suchman [46], defined his approach to evaluation), but it rather calls for linking theories, empirical knowledge and methodologies, or, as Chen and Rossi [4] put it: a linkage between basic and applied social science.‡

## CONCLUSION

The starting point for this article was the outline of a potential divergence between the mainstream approach to evaluation research and some of the contending approaches. Should this happen, evaluation research would be in danger of losing the confidence of decision makers and their continued support. Rather than taking sides in the emerging debate, the approach developed here proposes that the impressive bodies of knowledge which have been developed, often independently, in various fields in the social sciences, be viewed as a pool from which evaluators should learn to draw constructively. In this study we have sought to demonstrate how one such approach—integrated evaluation—may be assembled by linking selected components from the various approaches and by tailoring them to the needs of the decision-makers in a particular broad-aim social program.

## REFERENCES

1. J. S. Wholey, *Evaluation: Promise and Performance*. The Urban Institute, Washington, D.C. (1979).
2. E. R. House, *Evaluating with Validity*. Sage, Beverly Hills, California (1980).
3. G. Kress, G. Koehler, and J. F. Springer, Policy drift: evaluation of the California business enterprise program. In *Implementing Public Policy* (Edited by Dennis J. Palumbo and Marvin A. Harder), pp. 19–28. Lexington Books, Lexington, Massachusetts (1981).
4. Huey-Tsyh Chen and P. H. Rossi, The multi-goal, theory-driven approach to evaluation: a model linking basic and applied social science. *Social Forces* (1), 106–122 (September 1980).
5. E. B. Sharp, Models of implementation and policy evaluation: choice and its implications. In *Implementing Public Policy* (Edited by Dennis J. Palumbo and Marvin A. Harder), pp. 99–115. Lexington Books, Lexington, Massachusetts (1981).
6. M. Q. Patton, *Utilization-Focused Evaluation*. Sage, Beverly Hills, California (1978).
7. J. C. Thomas, "Patching up" evaluation designs: the case for process evaluations. In *Implementing Public Policy* (Edited by Dennis J. Palumbo and Marvin A. Harder), pp. 91–98. Lexington Books, Lexington, Massachusetts (1981).
8. D. J. Palumbo and M. A. Harder (eds.), *Implementing Public Policy*, Introduction, pp. IX–XVI. Lexington Books, Lexington, Massachusetts (1981).
9. M. Q. Patton, *Qualitative Evaluation Methods*. Sage, Beverly Hills, California (1980).
10. R. J. Madsen, Use of evaluation research methods in planning and policy contexts. *J. Plann. Ed. Res.* 2(2), 113–121 (1983).
11. R. Alterman, Implementation analysis in urban and regional planning: toward a research agenda. In *Planning Theory: Prospects for the 1980's* (Edited by Patsy Healey, Glen McDougall and Michael Thomas), Vol. 23, Chap. 15, pp. 225–245. Urban and Regional Planning Series, Pergamon Press (1982).

---

† The desire to trace and measure unintended effects caused Scriven [45] to propose what he termed "Goal-Free Evaluation." According to this approach, the actual effects of a program should be compared with a profile of "demonstrated needs". We agree with him when he says that evaluation is not supposed to reward the good intentions of the decision maker but rather to determine the program effects and evaluate them. However, we do not agree with the use of "needs" as being substitutions for goals; actually his needs are goals set by professionals to replace those which stemmed from the political process, and we object to this paternalistic dictation of goals.

‡ In their recent paper, Chen and Rossi [4] advocate a multi-goal theory-driven approach to evaluation. They provide strong arguments for using theory for identifying the program goals and developing the list of outputs to be measured by the evaluation study. However, they are still

---

only interested in assessing the program's effectiveness, in terms of ameliorating the problem for which it was designed. We argue that a program may have significant effects on groups and publics even if it is not effective in solving the social problems for which it was initially directed. Alternatively, it may have other effects in addition to contributing to the solution of the social problems for which it was conceived. The introduction of theory and empirical knowledge into the evaluation process may help in identifying such effects and measuring them.



12. R. Alterman, Implementation analysis: The contours of an emerging debate. *J. Plann. Ed. Res.* 2, (3), (Summer 1983).
13. M. Hill, *Planning for Multiple Objectives*. Regional Science Monograph No. 5, Philadelphia, Pennsylvania (1973).
14. N. Lichfield, P. Kettle and M. Whitbread, *Evaluation in the Planning Process*. Pergamon Press, London (1978).
15. G. Irvin, *Modern Cost-Benefit Methods*. Macmillan, London (1978).
16. D. M. McAllister, *Evaluation in Environmental Planning*. MIT Press, Cambridge, Massachusetts (1980).
17. M. Hill, A goals achievement matrix for the evaluation of alternative plans. *J. Am. Inst. Planners* 35(1), 19–28 (January 1968).
18. R. Rothman, Access versus environment. *Traffic Q.* 27, 111–131 (1973).
19. J. H. Schermer, Interest group impact assessment in transportation planning. *Traffic Q.* 29, 29–49 (1975).
20. D. Miller, Project location analysis using the goals-achievement method of evaluation. *J. Am. Plan. Ass.* 46, 195–208 (1980).
21. P. H. Rossi, H. E. Freeman and S. R. Wright, *Evaluation: A Systematic Approach*. Sage, Beverly Hills, California (1979).
22. F. Wedgwood-Oppenheim, D. Hael and B. Cobley, *An Exploratory Study in Strategic Monitoring*, Vol. 5, Part 1. Progress in Planning Series, Pergamon Press, Oxford (1976).
23. R. S. Weiss and M. Rein, The evaluation of broad-aim programs: experimental design, its difficulties, and an alternative. *Adm. Sci. Q.* 15(1), 17–109 (March 1970).
24. D. N. T. Perkins, Evaluating social interventions: a conceptual scheme. *Evaluation Q.* 1, 639–656 (November 1977).
25. P. Sabatier and D. Mazmanian, The implementation of public policy: a framework for analysis. *Policy Studies J.* 8(4), Summer (Special Issue No. 2), 538–559 (1980).
26. D. Van Meter and C. E. Van Horn, The policy implementation process: a conceptual framework. *Adm. Soc.* 6(4), 445–483 (February 1975).
27. R. S. Mountjoy and L. J. O'Toole, Towards a theory of policy implementation: an organizational perspective. *Public Adm. Rev.* 39(5), 465–476 (September–October 1975).
28. G. C. Edwards, *Implementing Public Policy*. Congressional Quarterly Press, Washington, D.C. (1981).
29. R. T. Nakamura and F. Smallwood, *The Politics of Policy Implementation*. St. Martin's Press, New York (1981).
30. S. M. Barrett and C. Fudge, *Policy and Action: Essays on the Implementation of Public Policy*. Methuen, London (1981).
31. T. H. Poister and A. H. Magoun, Local housing rehabilitation program: a low-effort evaluation. *Urban Affairs Papers* 1, 59–79 (Fall 1979).
32. E. Bardach, *The Implementation Game*. MIT Press, Cambridge, Massachusetts (1977).
33. J. L. Pressman and A. Wildavsky, *Implementation: How Great Expectations in Washington are Dashed in Oakland, etc.* University of California Press, Berkeley, California (1973).
34. E. Alexander, R. Alterman and H. Law-Yone, *Evaluating Plan Implementation: the National Statutory Planning System in Israel*. Progress in Planning Series, Pergamon Press 20, Pt. 2 (1983).
35. T. A. Goldman (ed.), *Cost-Effectiveness Analysis: New Approaches in Decision-Making*. Praeger, New York (1967).
36. H. Levin, Cost-effectiveness analysis in evaluation research. In *Handbook of Evaluation Research* (Edited by M. Guttentag and E. L. Struening), Vol. 2, pp. 89–124. Sage, Beverly Hills, California.
37. M. Thompson, *Benefit-Cost Analysis for Program Evaluation*. Sage, Beverly Hills, California (1980).
38. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. Rand McNally, Chicago (1963).
39. P. H. Rossi and W. Williams (eds.), *Evaluating Social Programs: Theory, Practice and Politics*. Seminar Press, New York (1972).
40. H. P. Hatry, R. E. Winnie and D. M. Fisk, *Practical Program Evaluation for State and Local Government Officials*. The Urban Institute, Washington, D.C. (1973).
41. H. W. Riecken and R. F. Boruch (eds.), *Social Experimentation*. Academic Press, New York (1974).
42. C. Selltitz, M. Hohoda, M. Deutsche and S. W. Cook, *Research Methods in Social Relations*. Henry Holt, Revised edition (1959).
43. J. L. Hochschild, *What's Fair-American Beliefs about Distributive Justice*. Harvard University Press, Cambridge, Massachusetts (1981).
44. D. Miller, *Social Justice*. Oxford University Press, London (1976).
45. M. Scriven, Pros and cons about goal-free evaluation. *J. Ed. Eval.* 3(4), Evaluation comments, 1–4 (December 1972).
46. E. Suchman, *Evaluative Research*. Russell Sage, New York. Takuya, Nakamoto (1982). Review essay: the changing field of evaluation. *J. Am. Plann. Ass.* 48(4), 515–517 (1967).