1-1-1990

Evaluating Programs

Ann B. Blalock, Editor

Washington State Employment Security Department

Follow this and additional works at: https://research.upjohn.org/up_bookchapters

---

Citation

Blalock, Ann B. 1990. "Evaluating Programs." In *Evaluating Social Programs at the State and Local Level: The JTPA Evaluation Design Project*, Ann Bonar Blalock, ed. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 3–40. <https://doi.org/10.17848/9780585243900.ch1>

This title is brought to you by the Upjohn Institute. For more information, please contact repository@upjohn.org.

1

Evaluating Programs

Ann Bonar Blalock
Washington Employment Security Department

Decisionmakers, planners, project staff, and [program] participants are increasingly skeptical of common sense and conventional wisdom as sufficient bases upon which to design social programs that will achieve their intended goals.

Peter Rossi and Howard Freeman
Evaluation: A Systematic Approach

Questions have always been raised about the value of tax-supported program strategies for resolving major problems in society. Resistance and inability to provide evidence of public value have, in the end, led to social unrest and sometimes active protest. This is not likely to change. Throughout history, those receiving and spending public money have had to guarantee that a reasonable tradeoff would be honored between their *rights* and *responsibilities*. Discretion and control have had to be balanced by an accounting of the public good produced, of the kinds and in the ways intended. Planned change has always required accountability.

There is a new element, however, in the historical exchange between public privileges and obligations. It involves the *form* accountability now takes. The scientific technology for developing evidence of value and popular support for using scientific approaches and methods have changed radically over the last three decades. Increased public education, breakthroughs in the automation of data processing and analysis, and growing interest in the role of applied social science research in weighing the risks and benefits of social initiatives have converged.

Evaluators can now perform statistical analyses of large amounts of data rapidly and relatively inexpensively. Evaluation research has become an established field within applied social research (Patton 1986), and the public's sophistication about the contribution evaluation re-

search can make to more objective judgments of social programs has grown significantly. Most elected officials, government professionals and others responsible for social programs now accept science-based evaluation as an integral part of public policymaking.

Evaluation is less well-understood, however, as a *pragmatic management tool* for making general improvements in programs and testing new ways to organize and deliver services for particular client groups. At the state and local level, evaluation has not been fully utilized as an opportunity to infuse the policy and planning process with better information, but there are cogent reasons why this condition has prevailed, and why the situation is changing.

While the value and legitimacy of evaluation was reinforced, the conservative social policies of the 1980s substantially changed the *location of responsibility* for making judgments of the value of social programs. Control and responsibility were transferred downward, from federal to state government. The premise for this change involved a mix of new and older political philosophies: new fiscal conservatism emerging in response to economic, energy, and environmental problems and postindustrial changes in American society; older conservatism favoring states' rights and opposing "big government"; and the more idiosyncratic social welfare policies of the Reagan administration, designed primarily to shrink federal involvement in social programs altogether (Palmer 1986). However, this shift in the locus of accountability occurred in the context of little prior state training or experience in performing the tasks required by the new evaluative role.

Although the federal evaluation effort throughout the 1970s was extensive, it did little to enhance state-level evaluation capability.¹ Collecting the data essential in carrying out federal evaluations required the cooperation of state and local program professionals, but these professionals were rarely brought into the effort as partners, and the information coming back to them from federal studies offered few direct informational benefits in adjusting policies or improving programs.

Although states were required to develop automated reporting systems responsive to federal program goals, few incentives existed to incorporate an evaluation capability within their design, or to develop in-

house evaluation expertise. When political changes required that states assume evaluation responsibilities, the infrastructure to support these activities was too often missing. Nevertheless, new provocative state and local oversight opportunities accompanied this unanticipated transfer of authority. Gradually recognizing these opportunities, states and local areas have begun to take a new interest in evaluation as an information tool.

The Purpose of the Book

This book is written for those at the state and local level who find themselves working within this new social welfare environment: those who develop program policies, oversee their translation into ongoing programs, and seek to judge their value.

Its purpose is to encourage them to view evaluation not simply as a new responsibility they now find themselves required to assume, but as a pragmatic policy and management tool that can directly inform their decisions and improve their programs. Its purpose is also to provide those given the responsibility for judging the value of state and local programs with practical guidance about how to develop or enhance their ability to evaluate, and offer them assistance in planning and carrying out program evaluations. Although these purposes are oriented primarily to the information needs and capabilities of those most involved with social programs at these levels, it is hoped that the book will be of interest and use to state and local elected officials, the evaluation research community, public interest and client advocacy groups, and ultimately the American taxpayer who makes social programs possible.

The new public-private decisionmaking partnerships, and the program administrators and staff who are now more responsible for seeing that programs meet their intent, have always wanted to be effective decisionmakers. Often they have not had sufficient in-depth information about their programs to feel comfortable about the decisions they must make. In an era that affords states and local areas vastly increased opportunities to think more imaginatively about defining accountability, evaluation is viewed in this book as a chance to obtain a better under-

standing of the much more complex and intriguing interplay of forces within programs, and between them and their environments. This understanding is necessary to the sound decisions practitioners want to make. Approaching evaluation as a new opportunity, states and local areas can use their increased autonomy to consider what information is essential to *them* in enhancing programs, particularly at the level where critical relationships between client and service provider occur.

Using Employment and Training Programs as a Case Example

Given the book's purpose, and its focus on the information needs of state and local program professionals, selecting an ongoing program as the key illustrative case throughout the book was considered a useful device for giving abstract concepts a richer and more practical meaning. Therefore, in each chapter the general ideas about evaluation, which are relevant to a range of social programs, are applied to a single case example.

The basic program for adults authorized in 1982 by the federal Job Training Partnership Act (JTPA) was selected as the illustration. JTPA exemplifies the contemporary trend in social programs, and its intricacies are well-known to the book's authors. The stimulus for writing the book was their development of a set of complementary evaluation guides for use by state and local JTPA council leaders and program practitioners.² Preparing the guides required a thorough knowledge of the legislative intent of JTPA and the actual operation of JTPA programs. This familiarity has permitted the authors to apply the concepts in the book to realistic program situations, drawing from actual employment and training program evaluations.

Frustrated by criticism of previous social program efforts to resolve unemployment and underemployment among the disadvantaged, and needing a bipartisan agreement on a new strategy, the architects of JTPA designed a program that made only incremental changes in the goals, services, and target groups of traditional employment and training programs, but substantially changed the philosophy and framework within which such programs were organized and implemented.

The proposed change agent in JTPA was the range of basic services common to previous approaches, with the major omission of public service employment and the restriction of training stipends to those most in need. The eligible client population remained the unemployed, the economically disadvantaged, the officially poor, and those with significant barriers to employment. The philosophy underlying JTPA was unique and anticipated future trends. This new initiative involved reduced funding and the decentralization of power and responsibility to public-private partnerships at the state and local level.

For the first time in the history of national public employment and training programs, and consistent with the strengthened role of the private sector, specific methods of quality control, i.e., federal *performance standards* that were to be met or exceeded by local service delivery areas (SDAs), were made an integral part of the legislation. State fiscal rewards and state sanctions in the form of mandatory technical assistance were incorporated into the legislation as reinforcements for local conformance with standards.

JTPA also required increased coordination with related programs, and encouraged new modes of organizing the provision of services, including performance-based subcontracting. Most relevant to this book, states and local areas were expected to move beyond *monitoring* program outcomes against a limited set of performance measures. The legislation mandated that states *evaluate* their programs as well. Seemingly novel at the time, many of JTPA's distinctive features became models for future publicly funded social programs. The reader will find it useful, in understanding the application of the book's central ideas to JTPA, to refer to the more detailed description of this program in the appendix.

The Framework of the Book

The *concept of evaluation* that provides the framework for this book is not an orthodox one. Despite the prolific literature that now characterizes the special field of evaluation research, most books and articles on program evaluation have been written primarily for researchers or oriented to those dealing with program issues at the national level. This

literature has only rarely addressed state and local evaluation needs and interests, or been directed to those making decisions about social programs at levels closer to the client.

Consistent with its purpose, this book describes a somewhat novel concept of evaluation that is specifically responsive to the information needs of state and local program decisionmakers. This perspective on evaluation recognizes the expanded opportunities inherent in their new oversight responsibilities, and respects the singular constraints and supports for evaluation at the state and local level. It views program evaluation holistically as both an art and a science.

Studying programs in real-life environments does not take place in a social vacuum. It is impossible to isolate research activities from other important influences. The organizational, political, and technical activities involved in producing information to inform decisions are interdependent. Giving attention to the scientific principles and methods involved in studying programs is critical. Restricting the book to research issues would ignore the larger environment of evaluation that resists or supports the application of research. Therefore, the book seeks to integrate the technical aspects of evaluation with its organizational and political realities.

The content of the book focuses on the state and local users of evaluation information, the conditions under which they must make program decisions, and the kinds of decisions they are expected to make at key points in a program's planning and funding cycle. The assumption is that the timeliness and relevance of evaluation information, as well as its accuracy and reliability, predict its level of use. This does not presume that the mere presence of these qualities is sufficient to carve a predominant role for evaluation in shaping policies or directing the fate of programs. A strong visible role for information production of this kind is mitigated by many other compelling agendas.

Nevertheless, there are key points in the policy and planning cycle where personal, bureaucratic, and political influences compete with less subjective sources of information to produce decisions through negotiation, exchange, and compromise. The point in this decisionmaking process at which the *worth* of a program strategy must be judged is

particularly vulnerable to the convergence of a variety of pressures, but it is precisely at this juncture where competent program evaluation can have its day in court.

Its case will not be heard, however, if evaluation information is not directly relevant to the content and context of the decisions being made, or if it reveals too little sensitivity to decisionmakers' information needs and technical sophistication. On the other hand, in the absence of empirically based usable information at this decision point, program adjustments will inevitably be based on more biased, and potentially misguided, considerations. To the extent that objective information can serve as one important influence among many other factors, state and local policies and the programs they generate will come closer to accurately defining and effectively resolving social problems.

In this context, the *comprehensive* concept of evaluation proposed in this book has a number of dimensions. Each is important, but the strength of the book is in its attempt to integrate these elements into an overall *evaluation strategy*. These dimensions can be expressed as interdependent sequential propositions.

1. Potential users of evaluations need to be identified and educated about the practical benefits of becoming *informed consumers of evaluation research* and *using it as a decisionmaking tool*.
2. State and local program councils and agencies need to develop an adequate *organizational capability to evaluate* their programs—to initiate, plan, and implement evaluations in-house or with the assistance of outside research consultants.
3. Evaluations should be planned with the *primary goal of improving programs*.
4. Evaluations will be of minimal utility in improving programs if evaluation sponsors fail to *seek answers to questions that are responsive to the interests and concerns of key decisionmakers*.
5. Relevant *insights from previous program evaluations* need to be considered in planning evaluations if the information from research studies is to contribute to program improvement and to the *accumulation of knowledge about social programs* from which decisionmakers can draw in the future.

10 Evaluating Social Problems

6. The most useful knowledge base for decisionmaking is one that *integrates information obtained by using comparable evaluation approaches and methodologies* across different programs.
7. To be of maximum use in improving programs and contributing to general knowledge production, comparable evaluations should yield *complementary information about (a) program implementation, and (b) the relationship between services and outcomes.*
8. This complementary information will be more accurate and usable if studies of implementation and impact are based on the use of *common definitions and measures of the major variables*, involve the *same historical cohort of clients*, reflect the *same program context and environment*, and represent the *same time period*.
9. If this complementary information is to be fully used, it must be presented to decisionmakers in *user-oriented, nontechnical form at optimal points in the program planning cycle.*
10. Comprehensive evaluations providing complementary information to decisionmakers *make the best use of new state and local oversight opportunities.*

The Meaning of Evaluation

In promoting this concept of evaluation, it is important to define what we mean by program evaluation. Like any complex system, the organizations involved in implementing social programs are expected to perform certain essential functions that assure that programs are continued, can be fine-tuned to increase their effectiveness, and can be adapted to changing client needs and circumstances.

In exchange for funds, these organizations are typically required to *review* program plans, *monitor* the way those plans are carried out, and *evaluate* the extent to which programs meet their overall intent. These obligations, tied to fiscal support, are usually authorized by legislation or the formal policies of governing bodies. The contemporary form these expectations take is drawn from American industrial concepts. Although it is infinitely more difficult to define “production process,” “consumer products,” and functions such as “quality control” and “research and

development” when dealing with human services, the same principles that have characterized the industrial process have been applied to social programs.

Certain consumers are selected as the key recipients of the goods produced, particular steps are taken in producing these goods, a certain level of consumption is maintained, and the consumer is affected by the product in some identifiable way. The characteristics of the process and products and the costs and benefits of the effort are observed and recorded. Marketing the product may be necessary, and research and development are accepted as a basis for change in the consumers targeted, the nature of the process, the product itself, the way it is introduced to the consumer, and the effect of the product on the recipient. Historically, these expectations have always attached to public expenditures in one form or another. We have simply made them more explicit and measurable and adapted them to the human service sector.

It is natural, then, that social programs meet such expectations and that certain oversight functions be established to assure quality and accountability. As the methods for carrying out these oversight functions have become more precise, automated, and reliant on scientific method, the distinctions among them have grown clearer and more consistent across different areas of social legislation. In this decade, review, monitoring, and evaluation are generally understood to be *distinct but related* accountability functions that have different purposes and require different expertise. Collectively, these assessments seek to *preserve the distinctive features of a program*.

The typical differences among these functions are suggested in chart 1.1. Based on these distinctions, evaluation is defined as the *systematic collection, analysis, and interpretation of information to answer questions about the efficiency and effectiveness of program implementation and impact, using the principles and methods of social science research* (Rossi and Freeman 1985).

When we study a program’s efficiency, we want information on how well resources are used to achieve program goals; i.e., are we achieving maximum social benefits for a given outlay of resources? When we investigate program effectiveness, we want to know the tradeoff between

Chart 1.1
Program Accountability Functions

Oversight Responsibilities	Purpose	Definition	Information Produced	Expertise Required
Review	To determine the extent to which program <i>plans</i> are consistent with the legislative intent of the program.	One-time <i>planning</i> assessments of programs, special services, service delivery, and/or coordination <i>activities</i> judged against program <i>plans</i>	Information about the extent to which <i>program activities</i> comply with those <i>planned</i> .	Professional <i>planning</i> expertise
Monitoring	To determine the level of compliance with <i>administrative policies and procedures</i> designed to interpret the legislative intent of the program.	Ongoing <i>managerial</i> assessments of selected aspects of the organizations operating programs, service delivery systems, target groups, and/or services and outcomes, judged against program <i>rules and regulations</i> .	Information about the extent to which programs comply with rules and regulations — in the way they are being <i>implemented</i> and the <i>outcomes</i> being experienced by those exposed to their services	Professional <i>managerial</i> expertise
Evaluation	To determine to what extent a program is achieving its broader legislative intent in terms of producing the expected <i>effects</i> on the <i>individuals or other entities for whom the program was created</i> , using the <i>means</i> proposed.	One-time <i>scientific</i> assessments of the adequacy of program implementation and/or impact judged against the program's <i>broader intent</i> .	Information about <i>relationships among different aspects of programs</i> : • among different components of program <i>implementation</i> • among <i>client characteristics</i>, <i>service interventions</i>, and <i>outcomes</i>. • between <i>implementation</i> and <i>outcomes</i>.	Social science <i>research</i> expertise

nonmonetary costs and benefits as well as between monetary costs and benefits. Evaluation focuses not only on what is happening in a program, but also why and how. The emphasis is on studying *relationships* among different aspects of a program (Weiss 1972).

In gathering information on compliance for reporting purposes, monitoring provides information that is useful and sometimes essential to an evaluation effort. However, it is not the purpose of monitoring to collect the full range of data on the way programs are organized and carried out or on the relationship between their interventions and effects, which is required to answer critical relationship questions.

An interesting example of the differences between monitoring and evaluation is the mandating of official performance standards to monitor the outcomes of programs. The use of standards is an effort to formalize the provision of evidence that a certain level of goal achievement is being attained. The Department of Labor developed a set of standards, measures for each of the standards, and a monitoring process to compare JTPA employment, income, and welfare reduction outcomes and program costs against quantitative goals. Compliance with these standards is determined using a narrow set of easily accessible, low cost, quantitative measures, such as the number of clients placed in jobs and their hourly wages.

It would seem that using performance standards is a form of program evaluation, or at least illustrates a gray area between monitoring and evaluation. However, the performance standards strategy is limited to a set of proxies for a much broader array of evidence of program efficiency and effectiveness. Although we may be making an important assessment of a program's effects by monitoring selected outcomes against carefully selected but incomplete standards, we may also obtain a skewed view of a program's value.

Compliance with standards does not tell us anything about relationships among a more extensive set of potential outcomes, between outcomes and the organization of the program, or between outcomes and the way services are delivered to the client. Most significantly, the level of compliance does not inform us about the program's role *vis-a-vis* these outcomes; i.e., was compliance responsible for the outcomes, or were these outcomes due to chance or other influences?

Monitoring information is not intended to help us look in depth at what is going on and explain it (Blalock 1982). Unfortunately, however, monitoring requirements tend to preoccupy state and local administrators with a restricted view of outcomes, distracting them from considering the magnitude, quality, and durability of program effects, and dampening their natural curiosity about the “black box” that holds the secrets of how and why certain outcomes may have occurred. Consequently, decisions about adjusting programs are often made without sufficient information. It is helpful, then, to think about evaluation as distinct from monitoring regarding the questions each is expected to answer (see chart 1.2).

The Practical Benefits of State and Local Evaluation

Practitioners have little difficulty appreciating the need for planning, managing, and monitoring, or understanding the prerequisites and benefits of these activities. But they sometimes question what social science research can do for them that collective experience and conventional wisdom cannot. Classic stereotypes about evaluation persist: it is too costly, risky, esoteric, inconclusive, too easily underutilized or misused.

Evaluation *does* cost. However, the largest program outlay, other than personnel and services, has characteristically been the start-up costs for designing and placing in operation automated *management information systems (MISs)* for required monitoring purposes. As the reliability of the information in these systems has improved, and their technical sophistication for recording, storing, extracting, and analyzing data has grown, interrelated MISs now afford a basic information bank for evaluation, which can be easily and inexpensively supplemented with additional data. It is inefficient, in fact, *not* to use these systems as an evaluation opportunity. By making optimal use of the monitoring data already being routinely gathered, in terms of data analysis opportunities, evaluation can expand resources for improving state and local programs.

It is true that evaluations may sometimes produce results that policymakers and administrators would rather not know. Some evaluations may lead to the conclusion that a program has made little or no difference

Chart 1.2
Monitoring vs. Evaluation

Monitoring Questions	Evaluation Questions
Is the program complying with the way administrators interpret its intent through rules and regulations? That is, is the program consistent with its intent regarding • How funds are to be spent • The selection of the target group • The way the program is to be organized • The services to be provided • The service delivery process to be used • The way outcomes are to be measured and reported Are the following in conformance with these rules and regulations: • The target group served • The services provided • Client outcomes • Program organization • Costs	Is the program meeting its broader legislative intent <i>efficiently in terms of cost, and effectively in terms of achieving program goals?</i> • What is the scope and range of outcomes, positive and negative? • Were the outcomes those intended for the people expected to benefit? • Were the outcomes significantly different for participants than for similar individuals not receiving the program's services, and in what way? • Did the implementation features that distinguish the program from previous strategies work as intended? • What influence did these organizational and service delivery features seem to have in producing the outcomes? • What was the tradeoff between program costs and benefits?

for some of the people exposed to it. However, such results are useful in modifying unrealistic assumptions about the effect a program's implementation and service strategies should have, about the groups for whom such strategies are to work best, or about the kinds of changes needed to make programs more effective. Disappointing results can also alert practitioners to influences that should have been studied but were not, or to developing better definitions and measures of the factors that *were* studied. It may be problems such as these that explain nonsignificant or negative results rather than the characteristics of the program. Awareness of such issues can lead to a better understanding of programs and to more informed changes in the allocation and targeting of program resources.

It is well-known that evaluation findings and conclusions drawn from them are misused more often than we would prefer, but policy and management decisions clearly suffer more from an *absence* of scientific efforts to obtain objective information. More often than we think, evaluation offers new chances to demonstrate successful program aspects rarely revealed by monitoring information. Important outcomes that do not lend themselves to numerical measurement and are critical to a program's success are frequently missed in the monitoring process.

On the other hand, evaluation should not be oversold. Science cannot establish "the truth" beyond a doubt. Scientific method only brings us closer to the truth than more subjective and undisciplined methods. Nevertheless, myriad, rich, warm anecdotal reports no longer convince taxpayers, program managers, or legislators that a program is worth its cost, or even that it works. Accurate empirical evidence of the value of both *ends* and *means* is the contemporary measure of the worth of social initiatives.

Therefore, the challenge for states and local program areas at the beginning of a new decade is to develop the deeper and more expansive knowledge they must have in order to make important short- and long-term program decisions: how to determine needs appropriately, use resources wisely, target services so that those most in need or likely to benefit will have access to them, plan and manage programs effectively, coordinate with related programs in a meaningful way, and test new ideas as a basis for innovation.

An impressive benefit of evaluation in meeting this challenge is the *reduction of bias* in the information made available for decisionmaking. In the simplest terms, a scientific approach is *a guide map for obtaining the least-biased information available* about what is really happening in programs. Bias-free information must be an essential part of the range of influences to which decisionmakers expose themselves in making judgments of program value and recommending change.

Administrators often take the position that evaluation has little impact on the decisionmaking process, suggesting it is used mainly to support foregone decisions. This is often true, but this conclusion neglects the role evaluation plays in *narrowing the range of choices available to decisionmakers to those likely to be the best ones*. The development of performance standards in employment and training programs is again a good example. The measures selected to “stand for” employment, earnings, and welfare-reduction outcomes were based on a substantial number of previously undertaken national evaluations that tested the predictive ability of optional measures of these traditional program effects (Barnow 1987). Those with greatest predictive capability were suggested as a basis for federal decisions on performance standards indices. Administrative choices were thus directed to the most accurate, least-biased measures. The ultimate decision about standards favored these measures, benefiting from the framework for decisionmaking provided by insights from past evaluations.³

Changes made in programs are inevitable, but not always desirable. With increased state and local control, and the new flexibility to shape the purpose and content of change, administrators have the power to facilitate the kinds of change that will directly benefit their programs. Therefore, it is critical that state and local decisionmakers have a pool of unbiased information to balance other pressures for change.

Reducing Bias

Information about social programs is hardly lacking. The issues are the reliability and validity of that information. The practical benefits of program evaluation are tied to the need for *objective* information for

decisionmaking. Important decisions about the future of programs are made on the basis of diverse information sources that vary in their approximation of the truth about a program's impact and the effectiveness of its implementation. Decisions are clearly more useful to society if they are informed by accurate information. Well-designed and competently conducted program evaluations that are responsive to state and local concerns are the most trustworthy source of information for making choices among alternative courses of action. An important purpose of evaluations is, therefore, to reduce *information bias*.

Biases can be viewed simply as factors other than those studied in an evaluation that may, or probably do, affect the answers we obtain to evaluation questions. The degree to which such answers may be biased is a function of the extent to which an evaluation plan can *control for* extraneous influences that may explain evaluation results as well as or better than those being studied. That is, ideally we want to eliminate competing explanations of our findings, so that we are assured that what we have learned applies specifically to the relationships we have selected for evaluation attention (Smith 1981).

Based on the principles and methods that characterize scientific method, the logical chronology of steps in the research process is designed to control for the effects of bias to the extent possible. This chronology describes methods that are within the control of the evaluator and methods that are fallback strategies to compensate for potential bias outside the evaluator's control. However, not even this systematic approach can eliminate *all* biases, since even in the most competent research many influences cannot be anticipated or measured. Knowledge about their potential effects may be lacking. It may be very difficult to quantify such influences for purposes of analysis. The cost of accessing such information may be prohibitive. Organizational or political constraints may exist. Nevertheless, conforming as much as possible to the classic steps in the research process brings us closer to the truth than any other method (Blalock 1982).

An overview of these steps is provided in chart 1.3, with the caveat that, in practice, the questions to be answered in this progression are highly interdependent and must often be dealt with simultaneously. To

provide the reader with a clearer idea of the *kinds* of biases that may emerge at *each point* in the research process, a hypothetical work/welfare demonstration project is used as an illustration in chart 1.4.

The Range of Evaluation Options

As we try to conform to the research process, we should appreciate the alternatives available for evaluating state and local programs. A variety of distinctions can be made among these research options, but we will concentrate on two: (1) the *issues* given primary attention, and (2) the characteristics of the *research designs* used to study these issues.

Issues of Interest in Evaluating Social Programs: Studying Program Implementation vs. Program Outcomes

Most major national evaluations, such as studies of programs under the Comprehensive Employment and Training Act (CETA) and the more recent series of evaluations of state work/welfare programs, have focused singularly on a program's outcomes (Mirengoff 1982; Gueron 1986). Recognition of the significance of studying program implementation has evolved slowly, even though practitioners have always had an intuitive awareness of its importance. To select among various evaluation possibilities appropriately, in a way that is responsive to different information purposes and evaluation circumstances, we need to understand the difference between *outcome evaluations* and studies of program implementation, or *process evaluations*.

The major impetus for the development of social programs lies in the political arena as much as in the realities of life in societies. The definition of certain social phenomena as "problematic" is shaped strongly by the parameters of public debate, media attention, and the political process. Having been assigned problem status, pressures build to develop public policies to ameliorate or resolve these conditions. Some policies are subsequently translated into action programs through legislation. These programs are essentially strategies for change—change in individuals, organizations, institutions and their environments. Their purpose and content are progressively reinterpreted by a

Chart 1.3
The Evaluation Chronology

Steps in the Evaluation Process	Questions Answered
Conceptualization and Measurement	
Review of Past Evaluations Identification of insights from past research relevant to the subject of interest Identification of the assumptions underlying this research	What reasonably reliable and valid information does past research provide that can help direct this evaluation? What assumptions did the evaluators make that can provide an effective rationale for selecting questions to be answered in this evaluation?
Conceptualization Identification of the important <i>issues</i> to study Identification of the major <i>variables</i> to investigate Identification of the most important <i>relationships</i> on which to focus Formulation of a manageable set of top priority <i>research questions</i> to be answered. Identification of the <i>variables to be controlled</i> in answering the research questions Specification of the assumptions made about what is most essential to study.	What program issues are most significant to study? What influences have the greatest impact in shaping these issues? What relationships among the major influences should be the subject of the evaluation? What are the most important questions to be answered about these issues, influences, and relationships? What influences — <i>other than those selected for study</i> — may explain the answers yielded by the evaluation, and therefore must be given attention? How can the focus of the evaluation be rationalized?
Measurement Definition of the <i>variables</i> , or influences to be studied. Development of <i>indices</i> for each of the variables Definition and development of indices of the <i>variables to be controlled</i> .	How will the major influences be defined in order to study them? Consistent with the definition of these variables, what kinds of <i>measures</i> (data elements) will be used to <i>represent</i> each influence in the evaluation? How will influences other than what is being studied be defined and measured, in order to control for their potential effects?

Chart 1.3 (Continued)
The Evaluation Chronology

Steps in the Evaluation Process	Questions Answered
Methodology	
Research Design Development of a research approach, and a set of sampling, data collection and data analysis methods for answering the research questions, given the data available	What kind of research <i>guide</i> , or <i>technical map</i> , will be used to direct the evaluation, so that it will yield the <i>least biased</i> information?
• Sampling Development of a sampling strategy for selecting the subjects of the evaluation: the sample of clients, and a comparison group (if used).	How should the individuals, groups or organizations to be studied be selected, in order to be able to generalize the results of the evaluation? How closely does the comparison group (if any) resemble the client group?
• Data Collection Development of a strategy for acquiring the information needed.	What is the most cost-effective way to collect a given quality and quantity of data? How should data on the measures selected be gathered, in order to obtain the most reliable information possible?
• Data Analysis The development of a strategy for analyzing the data collected, appropriate to the research design.	What statistical or other methods are most appropriate for analyzing the information obtained, in order to produce the most valid results possible?

Chart 1.3 (Continued)
The Evaluation Chronology

Steps in the Evaluation Process	Questions Answered
Use of Results	
Summarization of Findings The presentation of the most important evaluation results, given the research questions	What was learned about the issues, influences, and relationships of interest? What are the answers to the research questions?
Interpretation of Findings The drawing of conclusions from the findings, given the qualifications to these results.	What conclusions can be drawn from the findings, considering the following: • What potential sources of bias were not measured, or were outside the evaluator's control, that may limit evaluation conclusions? • How do the nature of the research questions and the characteristics of those studied affect the usefulness of the conclusions? • To what extent can conclusions be generalized beyond the subjects of this evaluation, and the context and environment of the program evaluated?
Advice about Improving the Program Evaluated The development of recommendations about improving program policy and operation.	Based on the findings and conclusions, and considering the biases left unmeasured or uncontrolled, as well as the extent to which results can be generalized, what recommendations can be made from the evaluation that could improve program policies and operation?

Chart 1.4
Potential Sources of Bias in Evaluating a New Welfare Reform Project

Steps in Evaluation Process	Examples of Potential Sources of Bias
Review of Past Evaluations	If past evaluations of work/welfare programs are underutilized, or if the analysis of previous work/welfare studies does not give enough consideration to the methodology used or the qualifications that must be made to the findings, the evaluation questions that drive the evaluation of the demonstration project may be based on inadequate or faulty assumptions and conclusions; therefore, the answers may lack utility or be misleading. For example, some of the early evaluations of "workfare" programs have produced highly questionable results, whereas recent experimental studies of work/welfare programs have yielded much more trustworthy insights.
Conceptualization	If the least important, less relevant, or wrong variables and relationships are selected for study, the usefulness and cost-effectiveness of evaluation activities may not justify the effort. For example, the better studies of work/welfare programs suggest that the marital and family characteristics of the unit to be benefited by a program, and service interventions involving health services, day care, and supportive counseling, are key variables to measure in work/welfare programs. However, many studies have persisted in looking only at the standard demographic variables and the effects of standard employment and training services, probably missing the more important explanations for outcomes.
Measurement	If the essential service interventions, outcomes, and/or implementation features that characterize the new welfare reform program — and are the influences of greatest interest in evaluating it — are not defined precisely and accurately, the results will lack validity and lose their value in improving the new program. For example, in studying the effect of the mix and sequence of services provided to a given client, it is important to develop a precise, easily understandable definition of what each service in the mix is, and how sequence is to be defined. For example, the service intervention titled "classroom training" has many potential meanings. Precisely what activities it is to involve in the reform program must be established, and the meaning of "sequenced services" must be specified, if the effects of receiving this and other services, in particular sequences, are to be accurately determined.
Research Design	One set of biases that are difficult to control for, but must be dealt with in designing an evaluation, are <i>selection biases</i> . These can be biases due to clients' self-selection into a program, or biases due to the way program staff carry out intake, appraisal, service planning, service assignment and referral responsibilities. At each of the latter decision points, biases due to certain client and staff characteristics, and certain service delivery practices can interfere with obtaining valid results. For example, among those eligible only a subset apply for a program. With the reform, the applicant group may over-represent individuals of a certain age, with certain kinds of motivation and prior skills. The evaluator must somehow compensate for self-selection biases if the information desired from the evaluation is to apply to the entire group of eligibles.

Chart 1.4 (Continued)
Potential Sources of Bias in Evaluating a New Welfare Reform Project

Steps in Evaluation Process	Examples of Potential Sources of Bias
Research Design (Continued)	Staff frequently assign services on the basis of their own sense of who needs and will benefit from particular options, irrespective of formal policies and procedures — or formal policies themselves may require nonrandom assignment procedures. In evaluating the reform, the evaluator must determine what kinds of service assignment biases are likely to be introduced, so compensatory strategies can be used to reduce these biases.
Sampling	If the sample of individuals selected to be studied in evaluating the reform is not representative of those about whom information is desired, the results of the evaluation will lack generalizability — the results will only apply to a subset of those participating in the program, and it may even be difficult to identify precisely the kinds of individuals to whom the information does apply. For example, if we want to know the postprogram outcomes of all participants, but the sample we draw overselects participants with particular characteristics, we cannot then conclude that the outcomes hold for all participants — they may only hold true for the most disadvantaged or the most skilled program applicants. If the evaluator wants to compare individuals experiencing the reform with similar individuals who are experiencing the older program, and the comparison group is selected in such a manner that it does not adequately “match” the reform-treated individuals, then differences between these groups may simply reflect differences in prior individual characteristics rather than differences due to the program. For example, the comparison group may overrepresent men or minorities, skewing the outcome information.
Data Collection	If needed information is not collected consistent with the data gathering process that is part of the research design, the reliability of this information is in question. For example, if follow-up information is to describe the postprogram outcomes of all participants in the reform, but information is gathered only on those who found jobs following the program, or if follow-up information is collected only on those easiest to contact, it will be very difficult to compensate for the biases created.
Data Analysis	Methods of analysis must be appropriate to the purpose of the evaluation, the kind of information desired from it, and the nature of the data gathered. If the methods used are not appropriate vis-a-vis these criteria, or if appropriate methods are not applied correctly, the validity of the results will be jeopardized. For example, “content analysis” may be the most appropriate method for obtaining information about staff attitudes toward clients or new personnel responsibilities in the new reform, if questionnaires do not use scaled responses. If such responses are subsequently reclassified into scales after-the-fact, in order to use statistical analysis techniques, unknown biases may be introduced.

Chart 1.4 (Continued)
Potential Sources of Bias in Evaluating a New Welfare Reform Project

Steps in Evaluation Process	Examples of Potential Sources of Bias
Summary of Findings	The answers do not exist in a vacuum. Their context is the entire evaluation process. In reporting them, it is necessary to anticipate the conditions that must be imposed upon their interpretation.
Interpretation of Findings	In drawing conclusions from evaluation findings, the sources, nature, and extent of bias must be considered. Some of the biases introduced may have been eliminated or minimized by studying certain aspects of program implementation and compensating through the use of statistical techniques. However, no evaluation is perfect. Conclusions, therefore, need to be qualified so that the user can make a judgment about the reliability, validity, and generalizability of the information produced, and the replicability of the evaluation in another program setting. For example, if the sample sizes used to study client subpopulations in the reform turn out to be too small to yield trustworthy information — for critical groups such as single parents with preschool vs. school-age children, for instance — then conclusions simply should not be drawn about these subgroups.
Advice about Improving the Evaluated Program	Researchers are not ordinarily expected to make policy or program recommendations based on their studies. However, social program evaluation is not solely a scientific effort. Its results are expected to lead to some kind of action. Just as the thrust of evaluation activities is to study the effects of planned change, evaluation information is to lead to positive action. Consequently the evaluation contract frequently requires that policy recommendations be made. The progression from <i>evaluation findings</i> to <i>conclusions</i> to <i>recommendations for action</i> is a sensitive one, potential biases presenting a challenge at each step. For example, the evaluation of the new welfare reform program may find that short-term welfare recipients benefited more from reform than chronic users. This does not necessarily justify leaping to the conclusion that the program does not work for long-term recipients of income subsidies, and the recommendation that services be targeted more exclusively to short-term welfare consumers. One would have to address issues such as the preprogram characteristics of these two groups and the differential effect of alternative service mixes and sequences on their outcomes.

succession of political and organizational actors with different perspectives and investments. Each program, however, can be understood in terms of a number of common attributes that reveal assumptions about (1) the nature of the problem the program is to address, (2) who and what should be changed, and (3) what types of changes are needed and how they are to be accomplished, i.e., the “theory of change” underlying the design of the program.

Each social program involves a variety of strategies for accomplishing change. The strategy that tends to receive the most public policy attention is the program’s formal *service interventions* directed to participants. These involve particular mixes and sequences of services—and sometimes subsidies—that respond to a particular problem and distinguish a given program from others. Less visible to the public, but just as important to study, is the program’s *implementation strategy*, or the organizational features of a program that are expected to provide an environment in which the service interventions can have maximum effect. This strategy represents an additional change agent, or set of *organizational interventions* that can influence the extent to which the desired changes are accomplished. These implementation characteristics are (1) the *structural* and *functional* aspects of the organizations responsible for a program, and (2) the nature of the *service delivery systems* operated by these organizations.

Service delivery systems expose the client to the program’s service interventions, facilitating a satisfactory exchange between program effort and the social benefits associated with change. The way this exchange occurs is greatly influenced by the organizational context in which it takes place. In turn, this context has a significant impact on the program experiences of clients and client outcomes. Therefore, process evaluations that focus on how effectively and efficiently a program’s organizational system works, and what influences it has on program outcomes, provide essential information. Studying implementation poses unique measurement dilemmas, however. Some aspects of organizational life cannot be defined in a form that lends itself to *statistical analysis*. Other methods can be used to summarize and draw inferences from such information, yielding useful insights about these lesser-known, but critical organizational interventions.

Outcome evaluations, on the other hand, focus more exclusively on relationships among the characteristics of clients entering a service delivery system that are thought to affect their outcomes, the program's service interventions targeted to these clients, and these clients' outcomes. Outcome evaluations typically seek information on the *net* effects of services on outcomes, requiring that the evaluator trade information scope for information accuracy.

Sometimes important measurement issues must be neglected in outcome studies, such as the precision of the definitions that distinguish one kind of service activity from another; the development of profiles that describe the most frequent configurations of services and sequences delivered; the capture of information on significant attributes of the service interventions, such as length, and the gathering of information on important characteristics of client outcomes, such as quality, magnitude, and duration. Furthermore, outcome evaluations have often been limited to the program's effects for participants at the point of program completion, neglecting longer-term impacts. And some outcome studies have ignored the program's impact on other key individuals and entities, such as employers, educators, or other social programs.

Nevertheless, studying outcomes is the major oversight responsibility in judging the value of social programs. If outcome evaluations are to provide information genuinely useful for policy, planning, and management purposes, the goal is to define and measure as inclusive a range of important interventions and outcomes as a feasible research plan permits. Process studies can then offer information regarding why and how the observed outcomes occur. *Comprehensive* program evaluations, therefore, will address both implementation and outcome issues, affording users *complementary* information about the different influences and effects that describe a particular program.

Again, using the work/welfare demonstration project as an example, chart 1.5 helps us sort out the different influences and effects studied in process vs. outcome evaluations.

Pressure to justify a program's costs by demonstrating its social utility has often encouraged resistance to conducting process and outcome evaluations simultaneously. Outcome evaluations represent an ineffi-

cient use of resources, however, if carried out in the absence of process evaluations that can help explain program effects. The goal of outcome evaluations—to estimate program impact—and the goal of process evaluations—to understand why such effects were produced and how service or implementation strategies can be modified to more fully and efficiently produce desired outcomes—are not in conflict. The achievement of both goals assures more comprehensive, complementary information for decisionmaking (Judd 1987).

The Characteristics of Evaluation Approaches and Methodologies

In seeking answers to complicated questions about program implementation and impact, the ultimate objective is to come as close as possible to establishing which influences are *causing* certain effects. Was it good case management or the formal service received that resulted in a better outcome? Untangling cause and effect is a difficult task. Science can give us accurate estimates but not certainty. And it can only offer greater precision if we follow scientific principles and methods very closely. Sometimes that is not possible. The questions lead us elsewhere, or there are organizational or political reasons why we cannot use a particular evaluation strategy. Not all of our evaluation options provide us with information on cause and effect. They may, however, offer extremely useful information of a different kind about the relationships in which we are interested.

One basis for differentiating among evaluation strategies is the extent to which different *research designs* produce information that explains *cause-effect* relationships, as compared with determining associations or *correlations* among influences (short of establishing cause and effect), or simply providing important clues or insights about the possible nature of such relationships. Following scientific principles and methods most closely, the *ideal experiment* gives us the most “unbiased” information about programs. In realistic social program settings, however, not even experimental designs can produce information miracles. Therefore, we need to appreciate the spectrum of design possibilities available to us. One way to describe this spectrum is found in chart 1.6.

Chart 1.5
Influences Studied in Process and Outcome Evaluations

Target Groups Expected to Receive Program Interventions						
The Program's Participants: <i>Public Assistance Recipients</i> Key Characteristics: • Age • Sex • Education • Marital status • Number of children • Work history • Income history • Frequency/duration of periods on public assistance			Other Entities Expected to Benefit from Program Other Welfare Programs • Other training and employment programs • Unemployment insurance program • Other educational programs • Health programs • Economic development programs			
Desired Outcomes of Program						
For Program Participants • Increased employability • Increased occupational skills • More durable employment • Sufficient earnings to reduce or eliminate the need for a welfare grant • Greater family stability/life quality			For Other Entities • Reduced costs • Reduced service demands • Greater coordination and less duplication of effort • Increased tax revenues • Increased job creation			
Potential Service Interventions Expected to Produce Desired Outcomes						
Interventions Targeted to Participants			Interventions Targeted to Other Entities			
Employment/ Training Services	Services in Support of Employment/Training	Subsidies Associated with Employment/ Training	To Employers	To Other Programs		
			Services	Subsidies	Discounts	
• Educational/occupational skills testing • Basic education • Preparation for work • Skills training • work experience, classroom and on-the-job training • Job search assistance and placement • Follow-up	• Day care • Health care • Individual/family counseling • Financial counseling	• Basic income support • Training-related stipends	• Worker recruitment and screening • Special supportive services to workers through program's case management/ counseling effort	• Wage subsidies • Training subsidies	• Income tax credits	• Increased efforts to coordinate related services • Increased efforts to create new jobs

Chart 1.5 (Continued)
Influences Studied In Process And Outcome Evaluations

Program Implementation Features Expected to Foster Desired Outcomes		
Characteristics of Welfare Reform Program As an Organizational System		Characteristics of the Welfare Reform Program's Service Delivery System
Structural Characteristics	Functional Characteristics	Policies and Practices
State Executive Committee	• Acquisition of funding • Needs assessment/resource allocation • Formulation of program policies • Coordination with related programs • Program accountability: plan review, program monitoring, program evaluation	• Outreach • Intake/eligibility determination • Appraisal/service planning • Service assignment • Case management • Supervision: in educational/training/employment sites • Referral to other programs/alternatives • Placement • Follow-up
State Advisory Committee	Provision of expert opinion to Executive Committee	
State Program Administration Structure	Central administration, planning, management, monitoring/reporting	
Local Program Management Structure	Local administration, planning, contracting, program operation, monitoring and reporting	
Public/Private Service Subcontractor Systems at the Community Level	Supervision of local service delivery systems and provision of services	
Community Councils	Provision of grassroots feedback from local clients and service providers to the local organizational	
Summary of Issues Investigated in Process vs. Outcome Evaluations		
Studying Program Implementation		Studying Program Outcomes
The focus is on relationships among the organizational elements of a program and the potential influence of these organizational factors on program outcomes.		The focus is on relationships between the interventions targeted to program participants and the outcomes of the program for participants and other entities.

The characteristics of evaluation alternatives relevant to information about cause and effect are not the only attributes that distinguish one evaluation option from another. One must consider other characteristics, such as the *rigor* of the research design and the *realism* and *generalizability* of the information different kinds of evaluations provide.⁴ Each alternative may vary in terms of these attributes, depending on the decisions made in developing the research design guiding an evaluation and the way this methodology is actually implemented. For example, for each type of design described in chart 1.6, there is a variety of methodological choices that differ in the extent to which they produce accurate information—information that realistically reflects a program’s natural state and can be generalized beyond the individuals or other entities studied.

Even the ideal experimental design has its own set of tradeoffs in this respect. It is the most rigorous in the scientific sense, in that it provides maximum control over certain kinds of bias. However, control is achieved by randomly assigning eligible clients to served vs. nonserved groups. This essentially changes the program’s expected implementation strategy, which calls for service assignment at the discretion of professional staff.

Even if staff do not subvert random assignment, and client advocacy groups do not demand that all eligibles be served, we still find ourselves asking, “How did random assignment *itself* affect program outcomes?” Or if, for a variety of reasons, the experimental sites for studying a program are not representative of the areas to which evaluation results are to be generalized, the information obtained will be of less value no matter how unbiased it is.

Experimental designs also place limits on the variables and relationships that can be studied, in order to institute control, which then leads us to wonder whether we have studied the most important influences and effects or missed something critical. In general, experimental designs are stronger with respect to rigor, or the ability to isolate certain of the effects of a program’s interventions, but they tend to be weaker on realism, and often on generalizability (Blalock 1964).

Quasi-experimental designs interfere much less with a program’s service delivery practices, but they produce information of varying

Chart 1.6
Research Designs

Type of Research Design	Primary Purpose	Examples
Nonexperimental Designs		
Exploratory	To <i>explore</i> relationships among implementation features, service interventions, outcomes, and/or the program's environment, in order to identify what factors operating in these relationships may be the most important ones to study, and to obtain insights about the nature of their influence for more systematic study.	<i>Participant-observation</i> and <i>case studies</i> ranging from informal to quite formalized studies.
Descriptive	To <i>describe</i> relationships among any or all of the above, focusing on influences and effects known or considered to be probable key factors, to determine the extent to which they are interrelated — i.e., associated statistically, or <i>correlated</i> — short of attempting to establish cause-effect.	Surveys or panel studies of program participants and other individuals to be changed by the program, ranging from relatively unstructured to highly sophisticated studies. An example is <i>descriptive gross outcome studies</i> that involve simple statistical analyses of program follow-up information.
Quasi-experimental	To <i>estimate cause-effect</i> relationships between a program's interventions and outcomes, comparing program participants' outcomes with a <i>statistically constructed comparison group</i> of individuals who have not received the programs' interventions.	<i>Differential gross outcome studies</i> , which compare the outcomes of particular subgroups of program participants who have been assigned different interventions, using one group as a <i>comparison group</i> for another, adjusting statistically for selection bias, and controlling for influences on the program's environment that could affect outcomes. <i>Net impact studies</i> , which compare the outcomes of program participants with those of a group of individuals as similar to these participants as possible and who have not received the interventions, statistically adjusting for selection bias, and controlling for the program's environment.
Experimental Designs	To <i>more precisely determine cause-effect</i> relationships between interventions and outcomes, comparing the outcomes of participants who received the program's interventions with those who have not received them, based on the <i>random assignment</i> of a pool of clients to a "treated" group and a "non-treated control group" prior to the introduction of the interventions	<i>Field experiments</i> , which compare the outcomes of "treated" vs "non-treated" groups that are <i>equivalent</i> in terms of preprogram characteristics that may affect outcomes, controlling statistically for influences in the program's environment.

accuracy depending on the methods used to control for bias. Again, the scope of an evaluation is necessarily limited, but there is greater flexibility in studying programs as they actually operate. If carefully developed, these designs can achieve the desired generalizability in terms of identifying program outcomes across diverse sites. Therefore, quasi-experimental designs are generally weaker in controlling bias than experimental designs, but frequently stronger with respect to realism and generalizability (Campbell 1963).

Exploratory and descriptive designs clearly do not offer as much control over biases as the two designs just discussed. Assuring that the information from surveys, panel studies, and case studies has been purged of subjective judgments is much more difficult. Within this constraint, however, the sampling, data collection, and data analysis methods incorporated within descriptive designs can vary from relatively unsophisticated to highly sophisticated.

Well-structured exploratory designs can also yield reliable information. Far more flexible in what can be studied, the more rigorous exploratory and descriptive designs can reflect the true complexities of programs in their natural settings better than other methodologies. If sampling strategies are adequate, information obtained using descriptive designs is very generalizable, competent survey research being a case in point. And these designs are appropriate for studying both implementation and outcomes (Davis 1967). However, they represent weaker methodologies in terms of rigor.

These tradeoffs among accuracy, realism, and generalizability need to be viewed positively, as providing a framework within which different compromises can be crafted in the context of other tradeoffs, particularly organizational and political ones. As we try to increase our proximity to the truth, we should appreciate the different tradeoffs and options available and make the best choice given the many demands and pressures surrounding that choice.

The Evaluation Approaches Emphasized in this Book

The three nonexperimental evaluation approaches selected for attention in this book represent only one set of choices. They were chosen

because (1) they respond to the organizational and political realities of ongoing social programs; (2) they provide the least-biased information short of imposing random assignment procedures on service assignment policies; (3) they represent an integrated and comprehensive approach to studying programs over their planning cycles; and (4) they are oriented to the practical decisionmaking needs of those responsible for the success of such programs.

Net Impact Evaluation

The approach used in determining the net impact of programs, or their “return on the public investment,” is based on the principle that this return cannot be estimated precisely without comparing the outcomes of participants with those of similar individuals who have not received the program’s interventions. Therefore, the selection and measurement of an appropriate comparison group are key tasks. In this context, cause-effect relationships between interventions and outcomes for men vs. women are studied, using a rigorous quasi-experimental design. The influence of the length of time a client is exposed to a particular service intervention and the effects of local program environments are considered, focusing on the magnitude and duration of the outcomes achieved. A comparison group strategy utilizes regularly collected administrative data and adjusts for potential nonequivalence in the “treated” vs. “nontreated” groups through the application of statistical techniques.

Because of the rigorous research design required to estimate net program impact, such evaluations are necessarily more limited in the number and kinds of questions they can answer. However, they afford the best means of establishing what exclusive contribution the program is making to the short-term and longer-term fiscal and programmatic outcomes observed, as compared with other potential influences that could explain these outcomes.

Gross Outcome Evaluation

The approach used in studying a program’s gross outcomes is a twofold strategy for gaining an understanding of a broader range of outcomes than those that can be studied in net impact evaluations and

acquiring more in-depth information about the intensity, quality, and durability of such outcomes for client subgroups receiving different program interventions. This approach can make optimal use of and build upon a program's automated information system.

The first strategy utilizes a descriptive research design to study relationships between service interventions and gross client outcomes, as well as gross outcomes for others to be benefited by the program. The second strategy, referred to as a *differential gross outcome analysis*, uses a quasi-experimental design to analyze potential cause-effect relationships between program services and outcomes, sorted by types of interventions and kinds of clients. This approach does not involve the use of a comparison group of nontreated individuals. It relies, instead, on using one treated group as a comparison for another. This strategy also involves a study of service assignment practices and uses statistical techniques to compensate for selection biases. In this sense, the approach combines careful outcome evaluation with a complementary study of particular aspects of program implementation.

Process Evaluation

The approach taken in studying program implementation applies basic *systems analysis* concepts to social programs that are viewed as *organizational systems* (Mintzberg 1979). Emphasis is on how efficiently and effectively the parts of this system interrelate to achieve system goals, and on the influence the articulation among parts has on program outcomes. In evaluating implementation, the system is classified into its component parts and their integration with one another is examined. However, no matter how well such a system may work internally, it is also shaped by its relationships within a larger environment, including its funders, resource allocators, auditors, other human service organizations with which it is expected to interact, and key actors in its home community. Therefore, these relationships are also subjected to scrutiny.

The assumption underlying this approach—using either an exploratory or descriptive research design—is that organizations and systems of interrelated organizations are goal-directed, and successful goal achievement is a function of how well organizational structures and functions

work together. Information about these important organizational linkages can help explain why programs providing similar interventions to comparable clients can produce quite different outcomes. More often than not, the success of a program in conforming to its intent is more dependent on organizational factors than on the influences typically given maximum attention in the public policy arena. Such information is essential in making useful program modifications (Williams 1976).

The Organization of the Book

The book is organized to express the particular concept of evaluation outlined earlier. This concept views evaluation as a set of complementary approaches that can offer practical information on program implementation and impact for state and local decisionmaking. Evaluation is also seen as requiring a special fiscal, organizational, and political capability, which can benefit the state and local agencies operating social programs.

Chapters 1 through 4 focus on research approaches and methods for carrying out net impact, gross outcome, and process evaluations, respectively. Each set of approaches and methods has its own unique explanatory advantages in providing information useful in improving programs.

Chapter 5 encourages the reader to treat the development of a user-focused evaluation capability as an opportunity to increase staff skills and expand program resources. The author discusses practical planning challenges posed by evaluations conducted at state and local levels, and suggests realistic alternatives for acquiring the funds, expertise, and other resources essential to an evaluation effort.

The final chapter builds on the issues discussed in chapter 5 by proposing ways to strengthen evaluation capability and the use of results, and concludes with a brief overview of the book's major ideas.

Each of the chapters utilizes the same basic format, whose purpose is to assist the reader in absorbing complex material by moving from the general to the specific. This progression begins with an *overview* of the chapter's main ideas, which are relevant to the evaluation of a broad variety of social programs and demonstration projects. This overview is

followed by an *application* of these ideas to a contemporary case example, JTPA. Each chapter concludes with a brief exploration of how the ideas in the chapter *relate to and complement* ideas in other chapters, and a *summary* of the main points to be retained by the reader.

In the three technical chapters covering research strategies, the overview in some cases gives the reader a sense of the state-of-the-art regarding the evaluation approach proposed. Where feasible, the application of principles and methods to the JTPA example are presented in a manner consistent with the steps in the research process outlined earlier. Specific illustrations within these chapters are related to the kinds of pragmatic program decisions made by policymakers, administrators, planners, managers, or professionals responsible for program oversight. There is also a commitment to expressing technical concepts in practical language.

In conclusion, it is important to emphasize that looking at the political, organizational, and technical aspects of evaluation recognizes their influence in producing objective information for decisions affecting program evolution. The information user, evaluation funder, policymaker, program professional, and evaluator must work together as a novel partnership to guarantee that the information product justifies its effort and cost. This perspective on program evaluation, stressed throughout the book, acknowledges the reality that research techniques are only one major part of a larger system of influences that shape the form and content of program decisions.

NOTES

1. The decade of the 1970s witnessed a series of large-scale national human service demonstration projects, followed by substantial evaluations of their effects. These were innovative projects designed to test alternative program strategies with particular client groups. Examples are the extensive Seattle/Denver Income Maintenance Experiments, the Carter administration welfare reform projects, the four-state Flexible Intergovernmental Grant Project, and CETA. At the same time, there were a number of net impact studies of more established programs, such as the Work Incentive Program and Head Start. Most evaluations concentrated on the relationship between service interventions and client outcomes, but this decade marked the beginning of interest in the influence of program implementation as well. Social R&D were at their zenith. (See Mirengoff, Barrow, and Gueron for examples of major evaluations in the 1970s and 1980s.)

38 Evaluating Social Problems

2. The volumes and their authors are listed in the appendix. Another series of evaluation guides that may be of interest is a set of three volumes for studying local programs under CETA, prepared for Region I, Employment and Training Administration, U.S. Department of Labor in 1978 (Sum et al. 1978). Sage Publications has also produced a series of nine monographs on basic research approaches and methods titled the *Program Evaluation Kit* (Herman 1987).

3. The literature review in chapter 2 is also an excellent source of information about the basis for developing performance standards indicators in employment and training.

4. The three italicized terms, and the discussion of tradeoffs among these characteristics of research, are based on an interesting article by Leslie Kish, a well-known survey research expert, titled "Representation, Randomization, and Control" (1975).

REFERENCES

- Barnow, Burt. "The Impact of CETA Programs on Earnings: A Review of the Literature," *Journal of Human Resources* 22:2 (Spring 1987).
- Blalock, Ann B., and Hubert M. Blalock, Jr. *Introduction to Social Research*. 2d ed. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- Blalock, Hubert M., Jr. *Causal Inferences in Nonexperimental Research*. Chapel Hill, NC: University of North Carolina Press, 1964.
- Campbell, Donald T., and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1963.
- Cook, Robert, and Lane Rawlins. "The Job Training Partnership Act: New Federalism in Transition," *Publius* (Summer 1985).
- Cordray, David S., and Mark W. Lipsey, eds. *Evaluation Studies Review Annual: Volume 11*. Newbury Park, CA: Sage, 1987.
- Davis, James A. *Elementary Survey Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1967.
- Gueron, Judith. *Work Initiatives for Welfare Recipients: Lessons from a Multi-State Experiment*. New York: Manpower Demonstration Research Corporation, 1986.
- Harrison, Michael I. *Diagnosing Organizations*. Newbury Park, CA: Sage, 1987.
- Herman, Joan L., ed. *Program Evaluation Kit*. 2d ed. Newbury Park, CA: Sage, 1987.
- Judd, Charles. "Combining Process and Outcome Evaluation," In *Multiple Methods in Program Evaluation*, edited by Melvin Mark and R. Lance Shotland. San Francisco: Jossey-Bass, 1987.
- Kish, Leslie. "Representation, Randomization, and Control," In *Quantitative Sociology in the Social Sciences*, edited by Hubert M. Blalock, Jr. New York: Academic Press, 1975.
- Lincoln, Yvonna S. *Organizational Theory and Inquiry*. Newbury Park, CA: Sage, 1985.
- Lindblom, Charles E. *The Policymaking Process*. Englewood Cliffs, NJ: Prentice-Hall, 1968.
- Mintzberg, Henry. *The Structure of Organizations*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- Mirengoff, William. *CETA: Accomplishments, Problems, Solutions*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1982.
- Palmer, John L., ed. *Perspectives on the Reagan Years*. Washington, D.C.: The Urban Institute, 1986.
- Patton, Michael Q. *Utilization-Focused Evaluation*. 2d ed. Newbury Park, CA: Sage, 1986.

- Pressman, J.L., and Aaron B. Wildavsky. *Implementation*. 2d ed. Berkeley, CA: University of California Press, 1979.
- Rossi, Peter H., and Howard E. Freeman. *Evaluation: A Systematic Approach*. 4th ed. Newbury Park, CA: Sage, 1989.
- Rossi, Peter H., and Walter Williams, eds. *Evaluating Social Programs*. New York: Seminar Press, 1972.
- Simon, Julian, and Paul Burstein. *Basic Research Methods in Social Science*. 2d ed. New York: Random House, 1987.
- Smith, H. W. *Strategies of Social Research*. Englewood Cliffs, NJ: Prentice-Hall, 1981.
- Suchman, Edward A. *Evaluative Research*. New York: Russell Sage Foundation, 1967.
- Sum, Andrew, Katherine Mazzeo, Francis McLaughlin, and Jeffrey Zornitsky. *Evaluating the Performance of Employment and Training Programs at the Local Level*. Boston: Region I, Employment and Training Administration, U.S. Department of Labor, Vol.I-III, 1978.
- Weiss, Carol H. *Evaluation Research*. Englewood Cliffs, NJ: Prentice-Hall, 1972.
- Wholey, J. *Evaluation and Effective Public Management*. Boston, MA: Little, Brown, 1983.
- Williams, Walter, and Richard F. Elmore, eds. *Social Program Implementation*. New York: Academic Press, 1976.