

1

An Overview of Program Evaluation

Chapter Outline

What Is Program Evaluation?

A Brief History of Evaluation

- Evaluation Research as a Social Science Activity

- The Boom Period in Evaluation Research

- Social Policy and Public Administration Movements

- Development of Policy and Public Administration Specialists

- The Evaluation Enterprise From the Great Society to the Present Day

The Defining Characteristics of Program Evaluation

- Application of Social Research Methods

- The Effectiveness of Social Programs

- Adapting Evaluation to the Political and Organizational Context

- Informing Social Action to Improve Social Conditions

Evaluation Research in Practice

- Evaluation and the Volatility of Social Programs

- Scientific Versus Pragmatic Evaluation Postures

- Diversity in Evaluation Outlooks and Approaches

Who Can Do Evaluations?

2 Evaluation

In its broadest meaning, to evaluate means to ascertain the worth of or to fix a value on some object. In this book, we use evaluation in a more restricted sense, as program evaluation or interchangeably as evaluation research, defined as a social science activity directed at collecting, analyzing, interpreting, and communicating information about the workings and effectiveness of social programs. Evaluations are conducted for a variety of practical reasons: to aid in decisions concerning whether programs should be continued, improved, expanded, or curtailed; to assess the utility of new programs and initiatives; to increase the effectiveness of program management and administration; and to satisfy the accountability requirements of program sponsors. Evaluations also may contribute to substantive and methodological social science knowledge.

Understanding evaluation as currently practiced requires some appreciation of its history, its distinguishing concepts and purposes, and the inherent tensions and challenges that shape its practice. Program evaluation represents an adaptation of social research methods to the task of studying social interventions so that sound judgments can be drawn about the social problems addressed, and the design, implementation, impact, and efficiency of programs that address those problems. Individual evaluation studies, and the cumulation of knowledge from many such studies, can make vital contributions to informed social actions aimed at improving the human condition.

Since antiquity organized efforts have been undertaken to describe, understand, and ameliorate defects in the human condition. This book is rooted in the tradition of scientific study of social problems—a tradition that has aspired to improve the quality of our physical and social environments and enhance our individual and collective well-being through the systematic creation and application of knowledge. Although the terms *program evaluation* and *evaluation research* are relatively recent inventions, the activities that we will consider under these rubrics are not. They can be traced to the very beginnings of modern science. Three centuries ago, as Cronbach and colleagues (1980) point out, Thomas Hobbes and his contemporaries tried to calculate numerical measures to assess social conditions and identify the causes of mortality, morbidity, and social disorganization.

Even social experiments, the most technically challenging form of contemporary evaluation research, are hardly a recent invention. One of the earliest “social experiments” took place in the 1700s when a British naval captain observed the lack of scurvy among sailors serving on the ships of Mediterranean countries where citrus fruit was part of the rations. Thereupon he made half his crew consume limes while the other half continued with their regular diet. The good captain probably did not know that he was evaluating a demonstration project nor did he likely have an explicit “program

theory” (a term we will discuss later), namely, that scurvy is a consequence of a vitamin C deficiency and that limes are rich in vitamin C. Nevertheless, the intervention worked and British seamen eventually were compelled to consume citrus fruit regularly, a practice that gave rise to the still-popular label *limeys*. Incidentally, it took about 50 years before the captain’s “social program” was widely adopted. Then, as now, diffusion and acceptance of evaluation findings did not come easily.

What Is Program Evaluation?

At various times, policymakers, funding organizations, planners, program managers, taxpayers, or program clientele need to distinguish worthwhile **social programs**¹ from ineffective ones and launch new programs or revise existing ones so as to achieve certain desirable results. To do so, they must obtain answers to questions such as the following:

- What are the nature and scope of the problem? Where is it located, whom does it affect, how many are affected, and how does the problem affect them?
- What is it about the problem or its effects that justifies new, expanded, or modified social programs?
- What feasible interventions are likely to significantly ameliorate the problem?
- What are the appropriate target populations for intervention?
- Is a particular intervention reaching its target population?
- Is the intervention being implemented well? Are the intended services being provided?
- Is the intervention effective in attaining the desired goals or benefits?
- Is the program cost reasonable in relation to its effectiveness and benefits?

Answers to such questions are necessary for local or specialized programs, such as job training in a small town, a new mathematics curriculum for elementary schools, or the outpatient services of a community mental health clinic, as well as for broad national or state programs in such areas as health care, welfare, and educational reform. Providing those answers is the work of persons in the program evaluation field. Evaluators use social research methods to study, appraise, and help improve social programs, including the soundness of the programs’ diagnoses of the social problems they address, the way the programs are conceptualized and implemented, the outcomes they achieve, and their efficiency. (Exhibit 1-A conveys the views of one feisty senator about the need for evaluation evidence on the effectiveness of programs.)

1. Terms in boldface are defined in the Key Concepts list at the end of the chapter and in the Glossary.

4 Evaluation

EXHIBIT 1-A

A Veteran
Policymaker
Wants to See the
Evaluation Results

But all the while we were taking on this large—and, as we can now say, hugely successful—effort [deficit reduction], we were constantly besieged by administration officials wanting us to *add* money for this social program or that social program. . . . My favorite in this miscellany was something called “family preservation,” yet another categorical aid program (there were a dozen in place already) which amounted to a dollop of social services and a press release for some subcommittee chairman. The program was to cost \$930 million over five years, starting at \$60 million in fiscal year 1994. For three decades I had been watching families come apart in our society; now I was being told by seemingly everyone on the new team that one more program would do the trick. . . . At the risk of indiscretion, let me include in the record at this point a letter I wrote on July 28, 1993, to Dr. Laura D’Andrea Tyson, then the distinguished chairman of the Council of Economic Advisors, regarding the Family Preservation program:

Dear Dr. Tyson:

You will recall that last Thursday when you so kindly joined us at a meeting of the Democratic Policy Committee you and I discussed the President’s family preservation proposal. You indicated how much he supports the measure. I assured you I, too, support it, but went on to ask what evidence was there that it would have any effect. You assured me there were such data. Just for fun, I asked for two citations.

The next day we received a fax from Sharon Glied of your staff with a number of citations and a paper, “Evaluating the Results,” that appears to have been written by Frank Farrow of the Center for the Study of Social Policy here in Washington and Harold Richman at the Chapin Hall Center at the University of Chicago. The paper is quite direct: “Solid proof that family preservation services can affect a state’s overall placement rates is still lacking.”

Just yesterday, the same Chapin Hall Center released an “Evaluation of the Illinois Family First Placement Prevention Program: Final Report.” This was a large scale study of the Illinois Family First initiative authorized by the Illinois Family Preservation Act of 1987. It was “designed to test effects of this program on out-of-home placement and other outcomes, such as subsequent child maltreatment.” Data on case and service characteristics were provided by Family First caseworkers on approximately 4,500 cases: approximately 1,600 families participated in the randomized experiment. The findings are clear enough.

(Continued)

EXHIBIT 1-A

(Continued)

Overall, the Family First placement prevention program results in a slight increase in placement rates (when data from all experimental sites are combined). This effect disappears once case and site variations are taken into account. In other words, there are either negative effects or no effects.

This is nothing new. Here is Peter Rossi's conclusion in his 1992 paper, "Assessing Family Preservation Programs." Evaluations conducted to date "do not form a sufficient basis upon which to firmly decide whether family preservation programs are either effective or not." May I say to you that there is nothing in the least surprising in either of these findings? From the mid-60s on this has been the repeated, I almost want to say consistent, pattern of evaluation studies. Either few effects or negative effects. Thus the negative income tax experiments of the 1970s appeared to produce an increase in family breakup.

This pattern of "counterintuitive" findings first appeared in the '60s. Greeley and Rossi, some of my work, and Coleman's. To this day I cannot decide whether we are dealing here with an artifact of methodology or a much larger and more intractable fact of social programs. In any event, by 1978 we had Rossi's Iron Law. To wit: "If there is any empirical law that is emerging from the past decade of widespread evaluation activity, it is that the expected value for any measured effect of a social program is zero."

I write you at such length for what I believe to be an important purpose. In the last six months I have been repeatedly impressed by the number of members of the Clinton administration who have assured me with great vigor that something or other is known in an area of social policy which, to the best of my understanding, is not known at all. This seems to me perilous. It is quite possible to live with uncertainty, with the possibility, even the likelihood that one is wrong. But beware of certainty where none exists. Ideological certainty easily degenerates into an insistence upon ignorance.

The great strength of political conservatives at this time (and for a generation) is that they are open to the thought that matters are complex. Liberals got into a reflexive pattern of denying this. I had hoped twelve years in the wilderness might have changed this; it may be it has only reinforced it. If this is so, current revival of liberalism will be brief and inconsequential.

Respectfully,

Senator Daniel Patrick Moynihan

SOURCE: Adapted, with permission, from D. P. Moynihan, *Miles to Go: A Personal History of Social Policy* (Cambridge, MA: Harvard University Press, 1996), pp. 47-49.

6 Evaluation

Although this text emphasizes the evaluation of social programs, especially human service programs, program evaluation is not restricted to that arena. The broad scope of program evaluation can be seen in the evaluations of the U.S. General Accounting Office (GAO), which have covered the procurement and testing of military hardware, quality control for drinking water, the maintenance of major highways, the use of hormones to stimulate growth in beef cattle, and other organized activities far afield from human services.

Indeed, the techniques described in this text are useful in virtually all spheres of activity in which issues are raised about the effectiveness of organized social action. For example, the mass communication and advertising industries use essentially the same approaches in developing media programs and marketing products. Commercial and industrial corporations evaluate the procedures they use in selecting, training, and promoting employees and organizing their workforces. Political candidates develop their campaigns by evaluating the voter appeal of different strategies. Consumer products are tested for performance, durability, and safety. Administrators in both the public and private sectors often assess the managerial, fiscal, and personnel practices of their organizations. This list of examples could be extended indefinitely.

These various applications of evaluation are distinguished primarily by the nature and goals of the endeavors being evaluated. In this text, we have chosen to emphasize the evaluation of social programs—programs designed to benefit the human condition—rather than efforts that have such purposes as increasing profits or amassing influence and power. This choice stems from a desire to concentrate on a particularly significant and active area of evaluation as well as from a practical need to limit the scope of the book. Note that throughout this book we use the terms *evaluation*, *program evaluation*, and *evaluation research* interchangeably.

To illustrate the evaluation of social programs more concretely, we offer below examples of social programs that have been evaluated under the sponsorship of local, state, and federal government agencies, international organizations, private foundations and philanthropies, and both nonprofit and for-profit associations and corporations.

- In several major cities in the United States, a large private foundation provided funding to establish community health centers in low-income areas. The centers were intended as an alternative way for residents to obtain ambulatory patient care that they could otherwise obtain only from hospital outpatient clinics and emergency rooms at great public cost. It was further hoped that by improving access to such care, the clinics might increase timely treatment and thus reduce the need for lengthy and expensive hospital care. Evaluation indicated that centers often were cost-effective in comparison to hospital clinics.

- Advocates of school vouchers initiated a privately funded program in New York City for poor families with children in the first three grades of disadvantaged public schools. Scholarships were offered to eligible families to go toward tuition costs in the private schools of their choice. Some 14,000 scholarship applications were received, and 1,500 successful candidates were chosen by random selection. The evaluation team took advantage of this mode of selection by treating the program as a randomized experiment in order to compare the educational outcomes among those students who received scholarships and moved to private schools with the outcomes among those students not selected to receive scholarships.
- In recent decades, the federal government has allowed states to modify their welfare programs provided that the changes were evaluated for their effects on clients and costs. Some states instituted strong work and job training requirements, others put time limits on benefits, and a few prohibited increases in benefits for children born while on the welfare rolls. Evaluation research showed that such policies were capable of reducing welfare rolls and increasing employment. Many of the program features studied were incorporated in the federal welfare reforms passed in 1996 (Personal Responsibility and Work Opportunity Reconciliation Act).
- Fully two-thirds of the world's rural children suffer mild to severe malnutrition, with serious consequences for their health, physical growth, and mental development. A major demonstration of the potential for improving children's health status and mental development by providing dietary supplements was undertaken in Central America. Pregnant women, lactating mothers, and children from birth through age 12 were provided with a daily high-protein, high-calorie food supplement. The evaluation results showed that children benefited by the program exhibited major gains in physical growth and modest increases in cognitive functioning.
- In an effort to increase worker satisfaction and product quality, a large manufacturing company reorganized its employees into independent work teams. Within the teams, workers assigned tasks, recommended productivity quotas to management, and voted on the distribution of bonuses for productivity and quality improvements. An evaluation of the program revealed that it reduced days absent from the job, turnover rates, and similar measures of employee inefficiency.

These examples illustrate the diversity of social interventions that have been systematically evaluated. However, all of them involve one particular evaluation activity: assessing the outcomes of programs. As we will discuss later, evaluation may also focus on the need for a program, its design, operation and service delivery, or efficiency.

8 Evaluation

A Brief History of Evaluation

Although its historical roots extend to the 17th century, widespread systematic evaluation research is a relatively modern 20th-century development. The application of social research methods to program evaluation coincides with the growth and refinement of the research methods themselves as well as with ideological, political, and demographic changes.

Evaluation Research as a Social Science Activity

The systematic evaluation of social programs first became commonplace in education and public health. Prior to World War I, for instance, the most significant efforts were directed at assessing literacy and occupational training programs and public health initiatives to reduce mortality and morbidity from infectious diseases. By the 1930s, social scientists were using rigorous research methods to assess social programs in a variety of areas (Freeman, 1977). Lewin's pioneering "action research" studies and Lippitt and White's work on democratic and authoritarian leadership, for example, were widely influential evaluative studies. The famous Western Electric experiments on worker productivity that contributed the term *Hawthorne effect* to the social science lexicon date from this time as well. (See Bernstein and Freeman, 1975, for a more extended discussion and Bulmer, 1982, Cronbach and Associates, 1980, and Madaus and Stufflebeam, 1989, for somewhat different historical perspectives.)

From such beginnings, applied social research grew at an accelerating pace, with a strong boost provided by its contributions during World War II. Stouffer and his associates worked with the U.S. Army to develop procedures for monitoring soldier morale and evaluate personnel policies and propaganda techniques, while the Office of War Information used sample surveys to monitor civilian morale (Stouffer et al., 1949). A host of smaller studies assessed the efficacy of price controls and media campaigns to modify American eating habits. Similar social science efforts were mounted in Britain and elsewhere around the world.

The Boom Period in Evaluation Research

Following World War II, numerous major federal and privately funded programs were launched to provide urban development and housing, technological and cultural education, occupational training, and preventive health activities. It was also during this time that federal agencies and private foundations made major commitments to international programs for family planning, health and nutrition, and rural development. Expenditures were very large and consequently were accompanied by demands for "knowledge of results."

By the end of the 1950s, program evaluation was commonplace. Social scientists engaged in assessments of delinquency prevention programs, psychotherapeutic and psychopharmacological treatments, public housing programs, educational activities, community organization initiatives, and numerous other initiatives. Studies were undertaken not only in the United States, Europe, and other industrialized countries but also in less developed nations. Increasingly, evaluation components were included in programs for family planning in Asia, nutrition and health care in Latin America, and agricultural and community development in Africa (Freeman, Rossi, and Wright, 1980; Levine, Solomon, and Hellstern, 1981). Expanding knowledge of the methods of social research, including sample surveys and advanced statistical procedures, and increased funding and administrative know-how, made possible large-scale, multisite evaluation studies.

During the 1960s, the number of articles and books about evaluation research grew dramatically. Hayes's (1959) monograph on evaluation research in less developed countries, Suchman's (1967) review of evaluation research methods, and Campbell's (1969) call for social experimentation are a few illustrations. In the United States, a key impetus for the spurt of interest in evaluation research was the federal war on poverty, initiated under Lyndon Johnson's presidency. By the late 1960s, evaluation research had become a growth industry.

In the early 1970s, evaluation research emerged as a distinct specialty field in the social sciences. A variety of books appeared, including the first texts (Rossi and Williams, 1972; Weiss, 1972), critiques of the methodological quality of evaluation studies (Bernstein and Freeman, 1975), and discussions of the organizational and structural constraints on evaluation research (Riecken and Boruch, 1974). The first journal in evaluation, *Evaluation Review*, was launched in 1976 by Sage Publications. Other journals followed in rapid succession, and today there are at least a dozen devoted primarily to evaluation. During this period, special sessions on evaluation studies at meetings of academic and practitioner groups became commonplace, and professional associations specifically for evaluation researchers were founded (see Exhibit 1-B for a listing of the major journals and professional organizations). By 1980, Cronbach and his associates were able to state that "evaluation has become the liveliest frontier of American social science" (pp. 12-13).

As evaluation research matured, a qualitative change occurred. In its early years, evaluation was shaped mainly by the interests of social researchers. In later stages, however, the consumers of evaluation research exercised a significant influence on the field. Evaluation is now sustained primarily by funding from policymakers, program planners, and administrators who use the findings and by the interests of the general public and the clients of the programs evaluated. Evaluation results may not make front-page headlines, but they are often matters of intense concern to informed citizens, program sponsors, and decisionmakers, and those whose lives are affected, directly or indirectly, by the programs at issue.

10 Evaluation

EXHIBIT 1-B

Major Evaluation
Journals and
Professional
Organizations

Journals devoted primarily to program and policy evaluation:

Evaluation Review: A Journal of Applied Social Research
(Sage Publications)

American Journal of Evaluation (JAI Press) (formerly *Evaluation Practice*, before 1998)

New Directions for Evaluation (Jossey-Bass)

Evaluation: The International Journal of Theory, Research, and Practice (Sage Publications)

Evaluation and Program Planning (Pergamon)

Journal of Policy Analysis and Management (John Wiley)

Canadian Journal of Program Evaluation (University of Calgary Press)

Evaluation Journal of Australasia (Australasian Evaluation Society)

Evaluation & the Health Professions (Sage Publications)

Educational Evaluation and Policy Analysis (American Educational Research Association)

Assessment and Evaluation in Higher Education (Carfax Publishing Ltd.)

Professional organizations for program and policy evaluators:

American Evaluation Association

(Web page: <http://www.eval.org/>)

Association for Public Policy Analysis and Management

(Web page:)

American Educational Research Association (Evaluation Division)

(Web page: <http://aera.net>)

Canadian Evaluation Association (Web page:

<http://www.unites.uqam.ca/ces/ces-sce.html>)

Australasian Evaluation Society

(Web page: <http://www.parklane.com.au/aes/>)

European Evaluation Society

(Web page: <http://www.europeanevaluation.org>)

UK Evaluation Society

(Web page: <http://www.evaluation.org.uk>)

German Evaluation Society

(Web page: <http://www.fal.de/tissen/geproval.htm>)

Italian Evaluation Society

(Web page: <http://www.valutazione.it/>)

Incorporation of the consumer perspective into evaluation research has moved the field beyond academic social science. Evaluation has now become a political and managerial activity that makes significant input into the complex mosaic from which emerge policy decisions and resources for starting, enlarging, changing, or sustaining programs to better the human condition. In this regard, evaluation

research must be seen as an integral part of the social policy and public administration movements.

Social Policy and Public Administration Movements

Social programs and the associated evaluation activities have emerged from the relatively recent transfer of responsibility for the nation's social and environmental conditions, and the quality of life of its citizens, to government bodies. As Bremner (1956) has described, before World War I, except for war veterans, the provision of human services was seen primarily as the obligation of individuals and voluntary associations. Poor people, physically and mentally disabled persons, and troubled families were the clients of local charities staffed mainly by volunteers drawn from the ranks of the more fortunate. Our image of these volunteers as wealthy matrons toting baskets of food and hand-me-down clothing to give to the poor and unfortunate is only somewhat exaggerated. Along with civic associations, charity hospitals, county and state asylums, locally supported public schools, state normal schools, and sectarian old-age homes, volunteers were the bulwark of our human service "system." Indeed, government—particularly the federal government—was comparatively small before the 1930s. The idea of annual federal expenditures of billions of dollars to provide medical care for the aged and the poor, for instance, would have been mind-boggling for the government official of the 1920s. More dollars for public education currently flow from Washington in a few months than were spent by the federal treasury for that purpose in the entire first decade of the 20th century.

Correspondingly, there was little demand for social and economic information. Even in the late 1930s, federal expenditures for social science research and statistics were only \$40-\$50 million, as compared to many magnitudes that amount today. Also, before the 1930s key government officials typically were selected without regard to objective competence criteria; indeed, there were few ways of objectively determining competence. The professional civil service was a fraction of the size it is today, most jobs did not require technical know-how, and formal training programs were not widely available.

All this began to change in the 1930s. Human services grew at a rapid pace with the advent of the Great Depression, and so, of course, did government. In part because of the unwieldiness that accompanied this accelerated growth, there was strong pressure to apply the concepts and techniques of so-called scientific management, which were well regarded in industry, to government programs and activities. These ideas first took hold in the Department of Defense and then diffused to other government organizations, including human service agencies. Concepts and procedures for planning, budgeting, quality control, and accountability, as well as later, more sophisticated notions of cost-benefit analysis and system modeling, became the order of the day in the human services area.

12 Evaluation

Development of Policy and Public Administration Specialists

During this same period, scholars with social science training began to investigate the political, organizational, and administrative decision making that took place in executive departments and other government agencies. In part, their interests were purely academic—they wanted to understand how government worked. However, individuals in leadership positions in government agencies, who were groping for ways to deal with their large staffs and full coffers of funds, recognized a critical need for orderly, explicit ways to handle their policy, administrative, program, and planning responsibilities. They found many of the concepts, techniques, and principles from economics, political science, psychology, and sociology useful. The study of the public sector thus grew into the largely applied research specialty that is now most commonly called “policy science” or “policy analysis.”

As government became increasingly complex and technical, its programs could no longer be adequately managed by people hired as intelligent generalists or because of their connections with political patrons, relatives, or friends. Most middle management jobs and many senior executive positions required specific substantive and technical skills, and those who filled them needed either training or extensive experience to do their work competently (see Exhibit 1-C). In response, many graduate schools of management, public health, and social work began programs to train students for government positions and more specialized schools, generally with “public administration” in their titles, were established or expanded.

The importance of evaluation is now widely acknowledged by those in political and administrative roles in government. Many federal agencies have their own evaluation units, as do a large number of their state counterparts. In addition, federal, state, and local agencies regularly contract for program evaluation with university researchers or research firms and consultants. Although evaluation research continues to have an academic side oriented toward training, methodology, theory, and study of the nature and effects of social programs, it is now generally practiced in a context of policy making, program management, or client advocacy. Thus, not only is its history intertwined with the social policy and public administration movements, but its practice typically occurs in the same political and organizational arenas as policy analysis and public administration.

The Evaluation Enterprise From the Great Society to the Present Day

Evaluation activities increased rapidly during the Kennedy and Johnson presidencies of the 1960s, when social programs undertaken under the banners of the War on Poverty and the Great Society provided extensive resources to deal with unemployment, crime, urban deterioration, access to medical care, and mental health treatment

EXHIBIT 1-CThe Rise of
Policy Analysis

The steady growth in the number, variety, complexity, and social importance of policy issues confronting government is making increasing intellectual demands on public officials and their staffs. What should be done about nuclear safety, teenage pregnancies, urban decline, rising hospital costs, unemployment among black youth, violence toward spouses and children, and the disposal of toxic wastes? Many of these subjects were not on the public agenda 20 years ago. They are priority issues now, and new ones of a similar character emerge virtually every year. For most elected and appointed officials and their staffs, such complicated and controversial questions are outside the scope of their judgment and previous experience. Yet the questions cannot be sidestepped; government executives are expected to deal with them responsibly and effectively.

To aid them in thinking about and deciding on such matters, public officials have been depending to an increasing extent on knowledge derived from research, policy analysis, program evaluations, and statistics to inform or buttress their views. More often than in the past, elected and appointed officials in the various branches and levels of government, from federal judges to town selectmen, are citing studies, official data, and expert opinion in at least partial justification for their actions. Their staffs, which have been increasing in size and responsibility in recent decades, include growing numbers of people trained in or familiar with analytic techniques to gather and evaluate information. Increasing amounts of research, analysis, and data gathering are being done.

Because the power to influence policy is widely shared in our system of government, public officials seeking to influence policy—to play the policy game well—must be persuasive. Because of the changing character of policy issues, it is probably harder to be persuasive than it used to be. Seniority, affability, and clever “wheeling and dealing” may be relatively less influential than being generally knowledgeable and tough-minded, having the ability to offer ideas and solutions that can attract a wide following, or having a reputation as a well-informed critic. Increasingly, officials from the president on down lose influence in policy debates when they cannot get their numbers right or when their ideas and arguments are successfully challenged by opposing experts. Indeed, thorough and detailed command of an issue or problem is often mandatory. Legislatures are requiring executives to be experts in the programs and issues under their jurisdiction. Judges are requiring detailed proof that administrative decisions are not arbitrary and capricious. Budget officials demand positive program evaluations. The public demands accountability. Thus the dynamic processes whereby our political system confronts social problems are perceptibly, if not dramatically, raising the standards of substantive and managerial competence in the performance of public responsibilities.

SOURCE: Adapted, with permission, from Laurence E. Lynn, Jr., *Designing Public Policy* (Santa Monica, CA: Scott, Foresman, 1980).

14 Evaluation

EXHIBIT 1-D

The 1960s Growth in Policy Analysis and Evaluation Research

The year 1965 was an important one in the evolution of “policy analysis and evaluation research” as an independent branch of study. Two developments at the federal government level—the War on Poverty-Great Society initiative and the Executive Order establishing the Planning-Programming-Budgeting (PPB) system—were of signal importance in this regard. Both offered standing, legitimacy, and financial support to scholars who would turn their skills and interests toward examining the efficiency with which public measures allocate resources, their impacts on individual behavior, their effectiveness in attaining the objectives for which they were designed, and their effects on the well-being of rich versus poor, minority versus majority, and North versus South.

The War on Poverty-Great Society developments initiated in 1965 represented a set of social interventions on an unprecedented scale. All impacted by them wanted to know if they were working, and who was being affected by them and how. Those with the skills to answer these questions found both financial support and an interested audience for their efforts. And the social science community responded. The same year saw government-wide adoption of the formal evaluation and analysis methods that had earlier been applied in Robert McNamara’s Defense Department in the Planning-Programming-Budgeting system. A presidential Executive Order gave employment and financial support to thousands who wished to apply their analytical skills to such efficiency, effectiveness, and equity questions.

SOURCE: Robert H. Haveman, “Policy Analysis and Evaluation Research After Twenty Years,” *Policy Studies Journal*, 1987, 16:191-218.

(see Exhibit 1-D). These programs were often hurriedly put into place, and a significant portion were poorly conceived, improperly implemented, or ineffectively administered. Findings of limited effectiveness and poor benefit-to-cost ratios prompted widespread reappraisal of the magnitude of effects that can be expected from social programs.

Partly as a consequence of the apparent ineffectiveness of many initiatives, along with increasing expenditures for social programs in the face of rising fiscal conservatism, the decade of the 1970s was marked by increasing resistance to the expansion of government programs (Freeman and Solomon, 1979). This, in turn, brought about a change in emphasis in the evaluation field. In particular, increased attention was given to assessing the expenditures of social programs in comparison to their benefits and to demonstrating fiscal accountability and effective management. In the process, many fiscal and political conservatives, often skeptical about social science, joined the advocates of social programs in pressing for the information that evaluations provide.

Beginning with the Reagan presidency in the 1980s, and continuing to the present day, domestic federal expenditures were curtailed in an attempt to control inflation and reduce the federal deficit. Many of the largest cutbacks targeted social programs. Many states and cities adopted a similar posture; indeed, some of the local and state reactions to their deteriorating economic situations were particularly severe. These developments were partly a consequence of the distrust, hostility, and political actions of community members dismayed with the painful bite of income and property taxes. As we have indicated, however, they were also influenced by disenchantment with the modest effects and poor implementation of many of the programs most ardently championed by public officials, planners, and politicians in prior decades.

As should be apparent, social programs and, consequently, the evaluation enterprise are powerfully shaped by changing times. Political perspectives during recent decades, not only in the United States but also in a number of Western European countries, have been especially concerned with the balance of benefits and costs for social programs. On the intellectual front, both conservative and liberal critiques of the Great Society programs have had an impact on the evaluation field. Although these criticisms were sometimes based more on ideology than evidence, they also have drawn on evaluation results in condemning social programs. The evaluation field has thus been thrust into the middle of contentious debates about the very concept of social intervention and faced with new challenges to demonstrate that any major program initiative can be effective.

As we move into the 21st century, national policy is dominated by fiscal conservatism, the devolution of responsibility to the states, and continuing skepticism about social programs. These trends have mixed implications for evaluation. On the one hand, revisions and reforms in social programs require evaluation if anything is to be learned about their fiscal and social impacts, and a number of major national evaluations are under way (Rossi, 2001). On the other hand, much of the responsibility for conducting evaluation has devolved to the states along with the programs. Despite a steady increase in the amount and quality of evaluation conducted at state and local levels, many states do not have the capability to undertake rigorous evaluation on a regular basis or the means to develop such capability.

Regardless of political trends, two points seem clear about the current environment for evaluation. First, restraints on resources will continue to require funders to choose the social problem areas on which to concentrate resources and the programs that should be given priority. Second, intensive scrutiny of existing programs will continue because of the pressure to curtail or dismantle those that do not demonstrate that they are effective and efficient. Moreover, both dissatisfaction with existing programs and shifts in political currents will result in new and modified programs that come forward with promises of being more effective and less costly. All these circumstances create an important role for evaluation research.

The Defining Characteristics of Program Evaluation

With the benefit of some historical context, we will now provide a fuller description of the nature of the program evaluation enterprise in contemporary context. We begin, as a proper textbook should, with a definition:

Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions. By elaborating on the components of this definition, we can highlight the major themes that we believe are integral to the practice of program evaluation.

Application of Social Research Methods

The concept of evaluation entails, on the one hand, a description of the performance of the entity being evaluated and, on the other, some standards or criteria for judging that performance (see Exhibit 1-E). It follows that a central task of the program evaluator is to construct a valid description of program performance in a form that permits incisive comparison with the applicable criteria. Failing to describe program performance with a reasonable degree of validity may distort a program's accomplishments, deny it credit for its successes, or overlook shortcomings for which it should be accountable. Moreover, an acceptable description must be detailed and precise. An unduly vague or equivocal description of program performance may make it impossible to determine with confidence whether the performance actually meets the appropriate standard.

Social research methods and the accompanying standards of methodological quality have been developed and refined over the years explicitly for the purpose of constructing sound factual descriptions of social phenomena. In particular, contemporary social science techniques of systematic observation, measurement, sampling, research design, and data analysis represent rather highly evolved procedures for producing valid, reliable, and precise characterizations of social behavior. Social research methods thus provide an especially appropriate approach to the task of describing program performance in ways that will be as credible and defensible as possible.

Regardless of the type of social intervention under study, therefore, evaluators will typically employ social research procedures for gathering, analyzing, and interpreting evidence about the performance of a program. In any event, commitment to the rules of social research is at the core of the perspective on evaluation presented in this book and is what is meant by the subtitle, *A Systematic Approach*. This is not to say, however, that we believe evaluation studies must follow some particular social research style or combination of styles, whether quantitative or qualitative, experimental or ethnographic, "positivistic" or "naturalistic." Nor does this commitment to the methods of social science mean that we

EXHIBIT 1-EThe Two Arms
of Evaluation

Evaluation is the process of determining the merit, worth, and value of things, and evaluations are the products of that process. . . . Evaluation is not the mere accumulation and summarizing of data that are clearly relevant for decision making, although there are still evaluation theorists who take that to be its definition. . . . In all contexts, gathering and analyzing the data that are needed for decision making—difficult though that often is—comprises only one of the two key components in evaluation; absent the other component, and absent a procedure for combining them, we simply lack anything that qualifies as an evaluation. *Consumer Reports* does not just test products and report the test scores; it (i) *rates or ranks* by (ii) *merit or cost-effectiveness*. To get to that kind of conclusion requires an input of something besides data, in the usual sense of that term. The second element is required to get to conclusions about merit or net benefits, and it consists of evaluative premises or standards. . . . A more straightforward approach is just to say that evaluation has two arms, only one of which is engaged in data-gathering. The other arm collects, clarifies, and verifies relevant values and standards.

SOURCE: Quoted, with permission, from Michael Scriven, *Evaluation Thesaurus*, 4th ed. (Newbury Park, CA: Sage, 1991), pp. 1, 4-5.

think current methods are beyond improvement. Evaluators must often innovate and improvise as they attempt to find ways to gather credible, defensible evidence about social programs. In fact, evaluators have been, and will likely continue to be, especially productive contributors to methodological development in applied social research.

Finally, this view does not imply that methodological quality is necessarily the most important aspect of an evaluation nor that only the highest technical standards, without compromise, are always appropriate for evaluation research. As Carol Weiss (1972) once observed, social programs are inherently inhospitable environments for research purposes. The circumstances surrounding specific programs, and the particular issues the evaluator is called on to address, frequently compel evaluators to compromise and adapt textbook methodological standards. The challenges to the evaluator are to match the research procedures to the evaluation questions and circumstances as well as possible and, whatever procedures are used, to apply them at the highest possible standard feasible to those questions and circumstances.

The Effectiveness of Social Programs

By definition, social programs are activities whose principal reason for existing is to “do good,” that is, to ameliorate a social problem or improve social conditions. It

18 Evaluation

follows that it is appropriate for the parties who invest in social programs to hold them accountable for their contribution to the social good. Correspondingly, any evaluation of such programs that is worthy of the name must evaluate—that is, judge—the quality of a program's performance as it relates to some aspect of its effectiveness in producing social benefits. More specifically, the evaluation of a program generally involves assessing one or more of five domains: (1) the need for the program, (2) the program's design, (3) its implementation and service delivery, (4) its impact, or outcomes, and (5) its efficiency. Subsequent chapters will address how evaluators make these assessments.

Adapting Evaluation to the Political and Organizational Context

Program evaluation is not a cut-and-dried activity like putting up a prefabricated house or checking a document with a word processor's spelling program. Rather, evaluators must tailor the initial evaluation plan to the particular program and its circumstances and then typically revise and modify their plan as needed. The specific form and scope of an evaluation depend primarily on its purposes and audience, the nature of the program being evaluated, and, not least, the political and organizational context within which the evaluation is conducted. Here we focus on the last of these factors, the context of the evaluation.

The evaluation plan is generally organized around the questions posed about the program by those who commission the evaluation, called the **evaluation sponsor**, and other pertinent **stakeholders**—individuals, groups, or organizations that have a significant interest in how well a program functions. These questions may be stipulated in very specific, fixed terms that allow little flexibility, as in a detailed contract for evaluation services. More often, however, the evaluator must negotiate with the evaluation sponsors and stakeholders to develop and refine the questions. Although these parties presumably know their own interests and purposes, they will not necessarily formulate their concerns in ways that the evaluator can use to structure an evaluation plan. For instance, the initial questions may be vague, overly general, or phrased in program jargon that must be translated for more general consumption. Occasionally, the evaluation questions put forward are essentially pro forma (e.g., is the program effective?) and have not emerged from careful reflection regarding the relevant issues. In such cases, the evaluator must probe thoroughly to determine what the question means to the evaluation sponsor and program stakeholders and why they are concerned about it.

Equally important are the reasons the questions about the program are being asked, especially the uses that will be made of the answers. An evaluation must provide information that addresses the issues that matter, develop that information in a way that is timely and meaningful for the decisionmakers, and communicate it in a form that is

EXHIBIT 1-FWhere Politics
and Evaluation Meet

Evaluation is a rational enterprise that takes place in a political context. Political considerations intrude in three major ways, and the evaluator who fails to recognize their presence is in for a series of shocks and frustrations:

First, the policies and programs with which evaluation deals are the creatures of political decisions. They were proposed, defined, debated, enacted, and funded through political processes, and in implementation they remain subject to pressures—both supportive and hostile—that arise out of the play of politics.

Second, because evaluation is undertaken in order to feed into decision making, its reports enter the political arena. There evaluative evidence of program outcomes has to compete for attention with other factors that carry weight in the political process.

Third, and perhaps least recognized, evaluation itself has a political stance. By its very nature, it makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeability of others, the legitimacy of program goals and program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation.

Knowing that political constraints and resistance exist is not a reason for abandoning evaluation research; rather, it is a precondition for usable evaluation research. Only when the evaluator has insight into the interests and motivations of other actors in the system, into the roles that he himself is consciously or inadvertently playing, the obstacles and opportunities that impinge upon the evaluative effort, and the limitations and possibilities for putting the results of evaluation to work—only with sensitivity to the politics of evaluation research—can the evaluator be as creative and strategically useful as he should be.

SOURCE: From Carol H. Weiss, "Where Politics and Evaluation Research Meet," *Evaluation Practice*, 1993, 14(1):94, where the original 1973 version was reprinted as one of the classics in the evaluation field.

usable for their purposes. For example, an evaluation might be designed one way if it is to provide information about the quality of service as feedback to the program director, who will use the results to incrementally improve the program, and quite another way if it is to provide the same kind of information to an external funder, who will use it to decide whether to renew the program's funding. In all cases, however, evaluators must design and conduct evaluations in a way that is sensitive to the political context (see Exhibit 1-F).

These assertions assume that an evaluation would not be undertaken unless there was an audience interested in receiving and, at least potentially, using the findings.

20 Evaluation

Unfortunately, sponsors sometimes commission evaluations with little intention of using the findings. For example, an evaluation may be conducted because it is mandated by program funders and then used only to demonstrate compliance with that requirement. Responsible evaluators try to avoid being drawn into such situations of “ritualistic” evaluation. An early step in planning an evaluation, therefore, is a thorough inquiry into the motivation of the evaluation sponsors, the intended purposes of the evaluation, and the uses to be made of the findings.

As a practical matter, an evaluation must also be tailored to the organizational makeup of the program. In designing the evaluation, the evaluator must take into account any number of organizational factors, such as the availability of administrative cooperation and support; the ways in which program files and data are kept and access permitted to them; the character of the services provided; and the nature, frequency, duration, and location of the contact between the program and its clients. In addition, once an evaluation is launched, it is common for changes and “in-flight” corrections to be required. Modifications, perhaps even compromises, may be necessary in the types, quantity, or quality of the data collected as a result of unanticipated practical or political obstacles, changes in the operation of the program, or shifts in the interests of the stakeholders.

Informing Social Action to Improve Social Conditions

We have emphasized that the role of evaluation is to provide answers to questions about a program that will be useful and will actually be used. This point is fundamental to evaluation—its purpose is to inform action. An evaluation, therefore, primarily addresses the audience (or, more accurately, audiences) with the potential to make decisions and take action on the basis of the evaluation results. The evaluation findings may assist such persons to make go/no-go decisions about specific program modifications or, perhaps, about initiation or continuation of entire programs. They may bear on political, practical, and resource considerations or make an impression on the views of individuals with influence. They may have direct effects on judgments of a program’s value as part of an oversight process that holds the program accountable for results. Or they may have indirect effects in shaping the way program issues are framed and the nature of the debate about them.

Program evaluations may also have social action purposes that are beyond those of the particular programs being evaluated. What is learned from an evaluation of one program, say, a drug use prevention program implemented at a particular high school, also tells us something about the whole category of similar programs, such as weight reduction or antismoking programs. Many of the parties involved with social interventions must make decisions and take action that relates to types of programs rather than

individual instances. A congressional committee may deliberate the merits of compensatory education programs, a state correctional department may consider instituting boot camps for juvenile offenders, or a philanthropic foundation may decide to promote and underwrite programs that provide visiting nurses to single mothers. The body of evaluation findings for programs of each of these types is very pertinent to decisions and social actions of similar sorts.

One important form of evaluation research is that which is conducted on demonstration programs, that is, social intervention projects designed and implemented explicitly to test the value of an innovative program concept. In such cases, the findings of the evaluation are significant because of what they reveal about the program concept and are used primarily by those involved in policy making and program development at levels broader than any one program. Another significant evaluation-related activity is the integration of the findings of multiple evaluations of a particular type of program into a synthesis that can inform policy making and program planning.

Evaluation thus informs social action, contributing information for planning and policy purposes, indicating whether certain innovative approaches to community problems are worth pursuing, or demonstrating the utility of some principle of professional practice. Evaluation research may even help shape our general understanding of how to bring about planned social change by testing the effects of certain broad forms of intervention. The common denominator in all evaluation research is that it is intended to be both useful and used, either directly and immediately or as an incremental contribution to a cumulative body of practical knowledge.

Evaluation Research in Practice

We have outlined the general considerations, purposes, and approaches that shape evaluation research and guide its application to any program situation. In practice, applying these concepts involves a balancing act between competing forces. Paramount among these is the inherent conflict between the requirements of systematic inquiry and data collection associated with evaluation research and the organizational imperatives of a social program devoted to delivering services and maintaining essential routine activities. The planning phase of evaluation and, especially, the data collection phase necessarily place unusual and sometimes unwelcome demands on a program's personnel and processes. Data collection, for instance, may require access to files, clients, staff, and facilities that disrupts normal program processes and distracts from or even compromises the service functions that are the program's primary obligation.

Every evaluation plan, therefore, must negotiate a middle way between optimizing the situation for research purposes and minimizing the disruption caused to normal

22 Evaluation

operations. We use the word *negotiate* quite deliberately here, because the best approach is for the evaluator to develop the evaluation plan in consultation with program personnel. If the needs and purposes of the evaluation are spelled out in detail before the research begins, and those personnel who will be affected (not just the administrators) are given an opportunity to react, make input, and otherwise help shape the data collection plan, the result is usually a more workable plan and better cooperation in the face of the inevitable strains the evaluation will place on the program's administrators and staff.

In addition to the conflict between the needs of the evaluation and the functions of the program, there are other inherent tensions in the practice of evaluation that warrant comment. Here we introduce a few of the more notable dilemmas the evaluator must confront: the incompatibility of a fixed evaluation plan with the volatility of social programs; the strain between a press for evaluations to be scientific, on the one hand, and pragmatic, on the other; and the competing approaches to evaluation offered up by a field of great diversity and little consensus.

Evaluation and the Volatility of Social Programs

One of the most challenging aspects of program evaluation is the continually changing decision-making milieu of the social programs that are evaluated. The resources, priorities, and relative influence of the various sponsors and stakeholders of social programs are dynamic and frequently change with shifts in political context and social trends. The 1996 welfare reform legislation, for example, has drastically altered the nature of income support for poor families. A program reconfiguration of this magnitude requires evaluations of family income support programs to be defined differently than in the past, with new outcomes and quite different program components at issue.

The priorities and responsibilities of the organizations implementing a program can also change in significant ways. For example, a school system relieved by the courts from forced school busing may lose interest in its programs to increase white students' acceptance of assignment to predominantly minority schools. Or unanticipated problems with the intervention may require modifying the program and, consequently, the evaluation plan as well. A program to reduce the absence rates of low-income high school students by providing comprehensive medical care might be thwarted if a large proportion of the eligible students refuse the services.

Somewhat ironically, preliminary findings from the evaluation itself may stimulate changes in a program that render the remainder of the evaluation plan obsolete. Consider, for example, a study of the impact of an alcohol treatment program that included six-month and one-year follow-ups of the clients. When the six-month follow-up revealed high rates of alcohol use among the treatment group, the program's staff markedly modified the intervention.

The evaluator must attempt to anticipate such program changes and prepare for them to the extent possible. More important, perhaps, is to match the evaluation to the program circumstances and prospects at the time the evaluation is planned. It would generally make little sense to design a rigorous assessment of the impact of a program that is under consideration for significant revision by relevant decisionmakers. Of equal importance, however, is the flexibility the evaluator brings to the evaluation task. Knowing the dynamic nature of programs, evaluators must be prepared to modify an evaluation if it becomes apparent that the original plan is no longer appropriate to the circumstances. This often involves difficult issues associated with the availability of resources for the evaluation, the time lines for producing results, and the relationships with the program administrators and evaluation sponsors, so it is not to be taken lightly. Social programs are not research laboratories, however, and evaluators must expect to be buffeted about by forces and events outside their control.

The contrast between the image of a research laboratory and the reality of social programs as places to conduct social research leads us directly to another of the inherent tensions in evaluation, that between a scientific and a pragmatic perspective on the process.

Scientific Versus Pragmatic Evaluation Postures

Perhaps the single most influential article in the evaluation field was written by the late Donald Campbell and published in 1969. This article outlined a perspective that Campbell advanced over several decades: Policy and program decisions should emerge from continual social experimentation that tests ways to improve social conditions. Campbell asserted that the technology of social research made it feasible to extend the experimental model to evaluation research to create an “experimenting society.” Although he tempered his position in later writing, it is fair to characterize him as fitting evaluation research into the scientific research paradigm (see Exhibit 1-G).

Campbell’s position was challenged by Lee Cronbach, another giant in the evaluation field. While acknowledging that scientific investigation and evaluation may use some of the same research procedures, Cronbach (1982) argued that the purpose of evaluation sharply differentiates it from scientific research. In his view, evaluation is more art than science and should be oriented toward meeting the needs of program decisionmakers and stakeholders. Whereas scientific studies strive principally to meet research standards, Cronbach thought evaluations should be dedicated to providing the maximally useful information that the political circumstances, program constraints, and available resources allow (see Exhibit 1-H).

One might be inclined to agree with both these views—that evaluations should meet high standards of scientific research and at the same time be dedicated to serving the information needs of program decisionmakers. The problem is that in practice

24 Evaluation

EXHIBIT 1-G

Reforms as Experiments

The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available.

SOURCE: Quoted from Donald Campbell, "Reforms as Experiments," *American Psychologist*, April 1969, 24:409.

EXHIBIT 1-H

Evaluators as Teachers

An evaluative study of a social program is justified to the extent that it facilitates the work of the polity. It therefore is to be judged primarily by its contribution to public thinking and to the quality of service provided subsequent to the evaluation. . . . An evaluation pays off to the extent that it offers ideas pertinent to pending actions and people think more clearly as a result. To enlighten, it must do more than amass good data. Timely communications—generally not "final" ones—should distribute information to the persons rightfully concerned, and those hearers should take the information into their thinking. To speak broadly, an evaluation ought *to inform and improve the operations of the social system*.

SOURCE: Quoted from Lee J. Cronbach and Associates, *Toward Reform of Program Evaluation* (San Francisco: Jossey-Bass, 1980), pp. 65-66.

these two goals often are not especially compatible. Conducting social research at a high scientific standard generally requires resources that exceed what is available for evaluation projects. These resources include time, because high-quality research cannot be done quickly, whereas program decisions often have to be made on short notice. They also include the funding needed for the expertise and level of effort required for high-quality scientific research. Moreover, research within the scientific framework may require structuring the inquiry in ways that do not mesh well with the perspectives of those who make decisions about the program. For example, specifying variables so that they are well defined and measurable under scientific standards may appear to trivialize what policymakers see as complex and dynamic facets of the program. Similarly, scientific standards for inferring causality, as when investigating program outcomes (was the program the cause of an observed change?), may require such elaborate experimental controls that what is studied is no longer the program's typical services, but some constrained version of uncertain relevance to the actual program.

Nor can one blithely dismiss scientific concerns in evaluation. Properly understood, the scientific approach is a very considered attempt to produce conclusions that are valid and credible. Even when an evaluation falls short of this ideal—as all do to some extent—science-based findings make an important contribution to a decision-making context that is otherwise rife with self-interested perceptions and assertions, ideological biases, and undocumented claims. But this statement, in turn, assumes that the evaluation conclusions meaningfully address the situation of concern to decisionmakers; if not, they may be praiseworthy for their validity and credibility, but still be irrelevant.

In practice, therefore, the evaluator must struggle to find a workable balance between the emphasis placed on procedures that ensure the validity of the findings and those that make the findings timely, meaningful, and useful to the consumers. Where that balance point should be will depend on the purposes of the evaluation, the nature of the program, and the political or decision-making context. In many cases, evaluations will justifiably be undertaken that are “good enough” for answering relevant policy and program questions even though program conditions or available resources prevent them from being the best possible designs from a scientific standpoint.

A further complication is that it is often unclear who the ultimate users of the evaluation will be and which of the potential users should be given priority in the design. An evaluation generally has various audiences, some with very immediate interests in particular aspects of the program under investigation, some with broader interests in the type of intervention the particular program represents, and others falling somewhere in between. Occasionally, the purposes and priority users of an evaluation are defined so clearly and explicitly in advance that the evaluator has relatively little difficulty in balancing scientific and pragmatic considerations. However, many evaluation situations are not so clear-cut. Evaluation may be routinely required as part of funding or contract arrangements with the presumption that it will be generally informative to a program’s managers, the evaluation sponsors, and other interested parties. Or it may evolve from a collaboration between a service agency with a need for information for management purposes and a researcher with broader interests in the type of intervention that a particular program provides. Under such circumstances, the trade-offs between utility for program decisionmakers and scientific rigor are such that it is rarely possible to design an evaluation that serves both interests well.

Some evaluation theorists champion **utilization of evaluation** as the overriding concern and advocate evaluation that is designed around the information needs of specific stakeholding consumers with whom the evaluator collaborates very closely (e.g., Patton, 1997). A contrary view is advanced by the authors of review articles in applied research journals who attempt to synthesize available research on the effectiveness of various interventions. These scholars generally deplore the poor methodological quality of evaluation studies and urge a higher standard. Some commentators want to have it

26 Evaluation

both ways and press the view that evaluations should strive to have utility to program stakeholders and contribute to cumulative knowledge about social intervention (Lipsey, 1997). Our outlook, for the didactic purposes of this book, is that all these options are defensible, but not necessarily equally defensible in any given evaluation situation. This, then, presents yet another issue for which the evaluator must make a judgment call and attempt to tailor the evaluation design to the particular purposes and circumstances.

Diversity in Evaluation Outlooks and Approaches

As the preceding discussion illustrates, the field of evaluation is diverse and contentious. The fundamental differences represented historically by Campbell and Cronbach represent but one instance of this diversity. Evaluation practitioners are drawn from a wide range of academic disciplines and professions with different orientations and methods, and this mix has contributed significantly to the multiplicity of perspectives. Other differences in outlook are related to the motivations of evaluators and the settings in which they work. The solo practitioner who undertakes short-term evaluations on contract with local agencies and the tenured professor with long-term foundation funding will likely have quite divergent views on their evaluation activities.

As the field of evaluation has matured and become institutionalized, interest has developed in explicating the different postures toward evaluation and the methods preferred by leaders in various “camps.” There is growing interest in identifying congruent elements among different perspectives to advance what is referred to as “evaluation theory” (Shadish, Cook, and Leviton, 1991). Advocates of the evaluation theory movement envision the development of a theory that will serve as the basis for decision making by evaluators as they proceed with their work (see Exhibit 1-I).

At present, we must acknowledge that evaluation is at least as much art as science, and perhaps should be and always will be. Inevitably, the evaluator’s task is to creatively weave together many competing concerns and objectives into a tapestry in which different viewers can find different messages. We recognize, too, the difficulty of teaching an art form, especially via the written word. Teaching evaluation is analogous to training physicians to be diagnosticians. Any intelligent person can be taught to understand the results from laboratory tests, but a doctor becomes an astute diagnostician only through practice, experience, and attention to the idiosyncrasies of each individual case. In this sense, learning from a text can provide only part of the knowledge needed to become a capable evaluator.

Who Can Do Evaluations?

Systematic evaluation is grounded in social science research techniques; hence, most evaluation specialists have had some social research training. Beyond that

EXHIBIT 1-IThe Ideal
Evaluation Theory

The ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront. It would (a) clarify the activities, processes, and goals of evaluation; (b) explicate relationships among evaluative activities and the processes and goals they facilitate; and (c) empirically test propositions to identify and address those that conflict with research and other critically appraised knowledge about evaluation.

SOURCE: Quoted from William R. Shadish, Thomas D. Cook, and Laura C. Leviton, *Foundations of Program Evaluation: Theories of Practice* (Newbury Park, CA: Sage, 1991), pp. 30-31.

commonality, evaluators are very diverse in their disciplinary backgrounds and professional training (see Exhibit 1-J). Ideally, every evaluator should be familiar with the full repertoire of social research methods. It is also critical for an evaluator to be knowledgeable about the target problem area the program addresses (e.g., crime, health, drug abuse) and about the findings from evaluations of previous interventions. This is necessary, first, for the evaluator to understand the issues and context with which the program deals and, second, for the evaluator to develop an appropriate evaluation plan that reflects the reality of the program and existing knowledge relevant to such programs.

At the most complex level, evaluation activities can be so technically complicated, sophisticated in conception, costly, and of such long duration that they require the dedicated participation of highly trained specialists at ease with the latest in social science theory, program knowledge, data collection methods, and statistical techniques. Such complex evaluations are usually conducted by specialized evaluation staffs. At the other extreme, there are many evaluation tasks that can be easily carried out by persons of modest expertise and experience.

It is the purpose of this book to provide an introduction to the field for those whose current positions, professional interests, or natural curiosity inspire them to want to learn how evaluations are conducted. Studying this text is, of course, only a start along the path to becoming an expert in evaluation—there is no substitute for experience. We also aim to provide persons responsible for administering and managing human services programs with sufficient understanding of evaluation concepts and methods to be able to judge for themselves what kinds of evaluations are appropriate to their programs and projects and to comprehend the results of those studies. In brief, we have tried to provide a text that is helpful to those who conduct evaluations, those who commission them, those who oversee evaluation staffs, and those who are consumers of evaluation research.

28 Evaluation

EXHIBIT 1-J

Diversity of the Members of the American Evaluation Association (in percentages)

Major Activity	2003	Organizational Setting	2003	Primary Discipline	1993
Evaluation	39	College or university	36	Education	22
Research	15	Private business	18	Psychology	18
Administration	10	Nonprofit organization	17	Evaluation	14
Teaching	8	Federal government agency	7	Statistical methods	10
Consulting	10	State/local government agency	6	Sociology	6
Student	6	School system	3	Economics and political science	6
Other	4	Other	6	Organizational development	3
Unknown	9	Unknown	8	Other and unknown	21

SOURCE: Data for 2003 based on 3,429 members, as reported by Susan Kistler, AEA (February 19, 2003). Data for 1993 adapted from *Evaluation Practice News* (October 1993); based on 2,045 AEA members as of June 1993.

Summary

- Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs. It draws on the techniques and concepts of social science disciplines and is intended to be useful for improving programs and informing social action aimed at ameliorating social problems.

- Modern evaluation research grew from pioneering efforts in the 1930s and burgeoned in the years after World War II as new methodologies were developed that could be applied to the rapidly growing social program arena. The social policy and public administration movements have contributed to the professionalization of the field and to the sophistication of the consumers of evaluation research.

- The need for program evaluation is undiminished in the current era and may even be expected to grow. Indeed, contemporary concern over the allocation of scarce resources makes it more essential than ever to evaluate the effectiveness of social interventions.

- Evaluation requires an accurate description of the program performance or characteristics at issue and assessment of them against relevant standards or criteria.

- Evaluation typically involves assessment of one or more of five program domains: (1) the need for the program, (2) the design of the program, (3) program implementation and service delivery, (4) program impact or outcomes, and (5) program efficiency. An evaluation must be tailored to the political and organizational context of the program being evaluated.

- In practice, program evaluation presents many challenges to the evaluator. Program circumstances and activities may change during the course of an evaluation, an appropriate balance must be found between scientific and pragmatic considerations in the evaluation design, and the wide diversity of perspectives and approaches in the evaluation field provide little firm guidance about how best to proceed with an evaluation.

- Most evaluators are trained either in one of the social sciences or in professional schools that offer applied social research courses. Highly specialized, technical, or complex evaluations may require specialized evaluation staffs. A basic knowledge of the evaluation field, however, is relevant not only to those who will perform evaluations but also to the consumers of evaluation research.

KEY CONCEPTS

Evaluation sponsor

The person, group, or organization that requests or requires the evaluation and provides the resources to conduct it.

Program evaluation

The use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action in ways that improve social conditions.

Social program; social intervention

An organized, planned, and usually ongoing effort designed to ameliorate a social problem or improve social conditions.

Social research methods

Procedures for studying social behavior devised by social scientists that are based on systematic observation and logical rules for drawing inferences from those observations.

30 Evaluation

Stakeholders

Individuals, groups, or organizations having a significant interest in how well a program functions, for instance, those with decision-making authority over the program, funders and sponsors, administrators and personnel, and clients or intended beneficiaries.

Utilization of evaluation

The use of the concepts and findings of an evaluation by decisionmakers and other stakeholders whether at the day-to-day management level or at broader funding or policy levels.