

21st Century Sociology

Evaluation Research

Contributors: MICHAEL QUINN PATTON
Editors: Clifton D. Bryant & Dennis L. Peck
Book Title: 21st Century Sociology
Chapter Title: "Evaluation Research"
Pub. Date: 2007
Access Date: December 10, 2014
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412916080
Online ISBN: 9781412939645
DOI: <http://dx.doi.org/10.4135/9781412939645.n98>
Print pages: II-360-II-369

©2007 SAGE Publications, Inc. All Rights Reserved.

This PDF has been generated from SAGE knowledge. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412939645.n98>

Chapter 98: Evaluation Research

[p. II-360 ↓]

To evaluate is to determine the value of something, that is, to determine its merit, worth, or significance. Evaluation research is the systematic application of scientific research procedures to inform evaluative judgments. Program evaluation, as one particular focus of this process, involves the systematic collection of empirical information about the activities, characteristics, and outcomes of programs to make judgments about the program's merit or worth, improve program effectiveness and/or inform decisions about future programming. *Merit* refers to the intrinsic value of a program, for example, how effective it is in meeting the needs of those it is intended to help. In schools, this means determining to what extent students are learning what they need to know. *Worth*, in contrast, refers to extrinsic value to those outside the program, for example, to the larger community or society. A welfare program that gets jobs for recipients has *merit* for those who move out of poverty and worth to society by reducing welfare costs. *Significance* involves determining the relevance and importance of evaluation research findings, for example, the extent one can confidently attribute observed outcomes to the program intervention.

This matter of defining evaluation is of considerable import because different evaluation approaches rest on different definitions. One traditional approach has been to define program evaluation as determining the extent to which a program attains its goals. However, as the practice of evaluation has evolved, program evaluation can and does involve examining much more than goal attainment. For example, evaluation research can include assessing the fidelity of program implementation, illuminating variations in program processes, searching for unanticipated consequences, and measuring actual needs in relation to immediate outcomes and long-term impacts. Measuring goal attainment, then, takes too narrow a focus to encompass the variety of ways evaluation research can be useful.

Evaluation researchers may use a variety of social science research methods to gather information, but they may also use management information system data, program

monitoring statistics, or other forms of systematic information that are not specifically research oriented. Evaluation research is a type of applied interdisciplinary social science and thereby differs fundamentally from basic research in the purpose of data collection and standards for judging quality. Basic scientific research is undertaken to discover new knowledge, test theories, establish truth, and generalize across time and space. Program evaluation is undertaken to inform decisions, clarify options, identify improvements, and provide information about programs and policies within contextual boundaries of time, place, values, and politics. The difference between basic research and evaluation research is the difference between conclusion-oriented and decision-oriented inquiry. *Conclusion-oriented basic research* aims to produce knowledge and test theory. *Decision-oriented evaluation research* informs and supports policy making, program decision making, and improvements in programs to increase effectiveness.

Diversity in Evaluation Research

Evaluation research is characterized by enormous diversity. From large-scale, long-term, international comparative [p. II-361 ↓] designs costing millions of dollars to small, short evaluations of a single component in a local agency, the variety is vast. Contrasts include internal versus external evaluations; outcomes versus process evaluation; experimental designs versus case studies; mandated accountability systems versus voluntary management efforts; academic studies versus informal action research by program staff; and published, polished evaluation reports versus oral briefings and discussions where no written report is ever generated. Then there are combinations and permutations of these contrasting approaches. To understand and appreciate this diversity it is helpful to understand the emergence and development of evaluation as a field of professional practice.

The Emergence of Evaluation as a Profession

Education has long been a primary target for evaluation, dominated by achievement testing. During the Cold War, after the Soviet Union launched Sputnik in 1957, calls for better educational assessments accompanied a critique born of fear that the education gap was even larger than the missile gap. Demand for better evaluations also accompanied the growing realization that, in years after the 1954 Supreme Court Brown decision requiring racial integration of schools, “separate and unequal” was still the norm rather than the exception. Passage of the U.S. Elementary and Secondary Education Act in 1965 contributed greatly to more comprehensive approaches to evaluation. The massive influx of federal money aimed at desegregation, innovation, compensatory education, greater equality of opportunity, teacher training, and higher student achievement was accompanied by calls for evaluation data to assess the effects on the nation's children. Policymakers were asking: To what extent did these changes really make an educational difference?

But education was only one arena for evaluation. Evaluation in the United States emerged in response to the demand to assess the federal projects spawned by the Great Society legislation of the 1960s. When the U.S. federal government began to take a major role in alleviating poverty, hunger, and joblessness during the Depression of the 1930s, the closest thing to evaluation was the employment of a few jobless academics to write program histories. It was not until the massive federal expenditures on an awesome assortment of programs during the 1960s and 1970s that accountability began to mean more than assessing staff sincerity or political head counts of opponents and proponents.

Great Society programs from the Office of Economic Opportunity were aimed at nothing less than the elimination of poverty. The creation of large-scale federal health programs, including community mental health centers, was coupled with a mandate for evaluation, often at a level of 1 to 3 percent of program budgets. Other major programs were created in housing, employment, services integration, community planning, urban renewal, welfare, criminal justice reform, community health care, and racial integration.

In the 1970s, these Great Society programs collided head on with the Vietnam War, rising inflation, increasing taxes, and the fall from glory of Keynesian economics.

Program evaluation as a distinct field of professional practice was born of two lessons from this period of largescale social experimentation and government intervention: first, the realization that there is not enough money to do all the things that citizens may want or demand and, second, that even if there were enough money, it takes more than money to solve complex human and social problems. As not everything can be done, there must be a basis for deciding which things are worth doing. Evaluation held the promise of helping determine where scarce resources could be best allocated for maximum impact.

While pragmatists turned to evaluation as a commonsensical way to figure out what works and is worth funding, visionaries were conceptualizing evaluation as the centerpiece of a new kind of society: *the experimenting society*. Donald T. Campbell gave voice to this vision in his 1971 address to the American Psychological Association as follows:

The experimenting society will be one which will vigorously try out proposed solutions to recurrent problems, which will make hard-headed and multidimensional evaluations of the outcomes, and which will move on to other alternatives when evaluation shows one reform to have been ineffective or harmful. We do not have such a society today. (Campbell 1991:223)

Early visions for evaluation, then, focused on evaluation's expected role in guiding funding decisions and differentiating the wheat from the chaff in federal programs. But as evaluations were implemented, a new role emerged: helping improve programs as they were implemented. The Great Society programs floundered on a host of problems: management weaknesses, cultural issues, and failure to take into account the enormously complex systems that contributed to poverty. Wanting to help is not the same as knowing how to help; likewise, having the money to help is not the same as knowing how to spend money in a helpful way. Many "War on Poverty" programs turned out to be patronizing, controlling, dependency generating, insulting, inadequate, misguided, overpromised, wasteful, and mismanaged. Evaluators were called on not

only to offer final judgments about the overall effectiveness of programs but also to gather process data and provide feedback to help solve programming problems along the way.

By the mid-1970s, interest in evaluation had grown to the point where two professional organizations were established: the academically oriented Evaluation Research Society and the practitioner-oriented Evaluation Network. In 1984, they merged to form the American Evaluation Association. By that time, interest in evaluation had become international with establishment of the Canadian Evaluation Society and the [p. II-362 ↓] Australasian Evaluation Society. In 1995, the international evaluation conference in Vancouver, Canada, included participation from new professional evaluation associations representing members of the European Evaluation Society. By 2004, there were over 40 national evaluation associations around the world and a new umbrella organization, the International Organization for Cooperation in Evaluation (Mertens 2005). Another organization formed to focus on evaluation in developing countries, the International Development Evaluation Association (IDEA). In 2005, in Toronto, Canada, participants in the international evaluation conference represented 55 countries.

Standards of Excellence for Evaluation

One major contribution of the professionalization of evaluation has been articulation of standards for evaluation. Before the field of evaluation identified and adopted its own standards, criteria for judging evaluations could scarcely be differentiated from criteria for judging research in the traditional social and behavioral sciences, namely, technical quality and methodological rigor. Methods decisions dominated the evaluation design process. Methodological rigor meant experimental designs, quantitative data, and sophisticated statistical analysis. Whether decision makers understood such analyses was not the researcher's problem. Validity, reliability, measurability, and generalizability were the dimensions that received the greatest attention in judging evaluation research proposals and reports.

It was in this context that evaluation standards were developed by a 17-member committee appointed by 12 professional organizations, including the American

Sociological Association, in deliberations that spanned five years with input from hundreds of practicing evaluation professionals. The standards published by the Joint Committee on Standards in 1981 dramatically reflected the ways in which the practice of evaluation had matured. The standards identified four areas of quality for judging evaluation research: utility, feasibility, propriety, and accuracy. Just prior to publication, Dan Stufflebeam (1980), Chair of the Committee, summarized the committee's work as follows:

I think it is interesting that the Joint Committee decided on that particular order [utility, feasibility, propriety, and accuracy]. Their rationale is that an evaluation should not be done at all if there is no prospect for its being useful to some audience. Second, it should not be done if it is not feasible to conduct it in political terms, or practicality terms, or cost effectiveness terms. Third, they do not think it should be done if we cannot demonstrate that it will be conducted fairly and ethically. Finally, if we can demonstrate that an evaluation will have utility, will be feasible and will be proper in its conduct, then they said we could turn to the difficult matters of the technical adequacy of the evaluation. (P. 90)

In 1994 and again in 2006 (forthcoming), revised standards were published following extensive reviews spanning several years (<http://www.wmich.edu/evalctr/jc/>). While some changes were made in the 30 individual standards, the overarching framework of four primary criteria remained unchanged: utility, feasibility, propriety, and accuracy. In particular, the standards made the criterion of use ascendant and primary. No matter how technically rigorous an evaluation research study may be, by the criteria of the standards, if the findings from an evaluation are not used, it is an inadequate evaluation.

Evaluation Use and the Sociology of Knowledge

The use of evaluation research can be viewed as a special application of the sociology of knowledge. The question of evaluation use became for evaluation professionals what sociologist C. Wright Mills (1959) called a critical public issue:

Issues have to do with matters that transcend these local environments of the individual and the range of his inner life. They have to do with the organization of many such milieux into the institutions of an historical society as a whole.... An issue, in fact, often involves a crisis in institutional arrangements. (Pp. 8–9)

The challenge of using evaluation in appropriate and meaningful ways represents just such a crisis in institutional arrangements. How evaluations are used affects the spending of billions of dollars to fight problems of poverty, disease, ignorance, joblessness, mental anguish, crime, hunger, and inequality. The issues include determining how programs aimed at combating societal ills are to be judged, how to distinguish effective from ineffective programs, how evaluations can be conducted in ways that lead to use, and how evaluation researchers avoid producing reports that gather dust on bookshelves, unread and unused. Those are the issues the utilization literature in evaluation address and that sociology of knowledge informs.

The issue of use has emerged at the interface between science and action, between knowing and doing, and is therefore a problem of applied sociology. Evaluation use raises fundamental questions about human rationality, decision making, and knowledge applied to creation of a better world.

The challenge of evaluation use epitomizes the more general challenge of knowledge use in contemporary society. Technology in the contemporary epoch, variously called The Information Age, The Communications Age, or The Knowledge Age, has developed the capacity to generate, store, retrieve, transmit, and instantaneously [p. II-363 ↓] communicate information and knowledge. Our problem is keeping up with, sorting

out, absorbing, and *using* information. Our technological capacity for gathering and computerizing information now far exceeds our human ability to process and make sense out of it all. We're constantly faced with deciding what's worth knowing versus what to ignore.

Getting people to use what is known has become a critical concern across the different knowledge sectors of society. A major specialty in medicine (compliance research) is dedicated to understanding why so many people fail to follow their doctor's orders. Common problems of information use underlie trying to get people to use seat belts, quit smoking, begin exercising, eat properly, and pay attention to evaluation findings. In the fields of nutrition, energy conservation, education, criminal justice, financial investment, human services, corporate management, international development—the list could go on and on—a central problem, often *the* central problem, is getting people to apply what is already known.

These examples of the challenges of putting knowledge to use set a general context for an applied sociology of knowledge approach to evaluation use: narrowing the gap between generating evaluation findings and actually using those findings for program decision making and improvement.

Evaluation and Rationality

Edward Suchman (1967) began his classic text on evaluation research with Hans Zetterberg's observation that “one of the most appealing ideas of our century is the notion that science can be put to work to provide solutions to social problems” (p. 1). Social and behavioral science embodied the hope of finally applying human rationality to the improvement of society. In 1961, Harvard-educated President John F. Kennedy welcomed scientists to the White House as never before. Scientific perspectives were taken into account in the writing of new social legislation. Economists, historians, psychologists, political scientists, and sociologists were all welcomed into the public arena to share in the reshaping of modern postindustrial society. They dreamed of and worked for a new order of rationality in government—a rationality undergirded by social scientists who, if not philosopher-kings themselves, were at least ministers to

philosopher-kings. Sociologist Carol Weiss (1974) has captured the optimism of that period as follows:

There was much hoopla about the rationality that social science would bring to the untidy world of government. It would provide hard data for planning... and give cause-and-effect theories for policy making, so that statesmen would know which variables to alter in order to effect the desired outcomes. It would bring to the assessment of alternative policies a knowledge of relative costs and benefits so that decision-makers could select the options with the highest payoff. And once policies were in operation, it would provide objective evaluation of their effectiveness so that necessary modifications could be made to improve performance. (P. 4)

By the end of the 1960s, it was becoming clear that evaluations of “Great Society” social programs were largely ignored or politicized. The utopian hopes for a scientific and rational society had somehow failed to be realized. The landing of the first human on the moon came and went, but poverty persisted despite the 1960's “War” on it—and research was still not being used as the basis for government decision making. Producing data is one thing, but putting such data to use is quite another matter. In the final analysis, the test of the effectiveness of outcome data is its impact on implemented policy. By this standard, there are significant questions about the number of successful evaluation studies for it has proved difficult to document many instances where evaluation research has had a direct effect on policy even when it has been specifically commissioned by government.

Nor is the challenge only one of increasing use. A parallel issue is that of misuse of findings. Evaluators must attend to *appropriate* use, not just amount of use. Results from poorly conceived studies have frequently been given wide publicity and findings from good studies have been improperly used. The field faces a dual challenge then: supporting and enhancing appropriate uses while also working to eliminate improper uses.

Diffusion of Innovations

The diffusion of innovations literature central to rural sociology has examined and attempted to explain the characteristics of innovations that affect adoption and dissemination (Rogers 1962; Rogers and Svenning 1969; Rogers and Shoemaker 1971). This was the framework that informed some early empirical work on evaluation use, for example, an inquiry that was the basis for the first edition of *Utilization-Focused Evaluation* (Patton 1978). That work was grounded in case studies of evaluations to find out what characteristics were associated with use (a form of adoption from a diffusion of innovations perspective). A related field in organizational sociology focuses on the characteristics of innovative organizations.

Power and Evaluation Use

Another root sociological influence in understanding evaluation use has been theories of power that illuminate what evaluation offers stakeholders and intended users. Examining evaluation from this perspective provides a basis for understanding how knowledge is power, which led to the following premise: Use of evaluation will occur [p. II-364 ↓] in direct proportion to its power-enhancing capability. Power-enhancing capability is determined as follows: The power of evaluation varies directly with the degree to which the findings reduce the uncertainty of action for specific stakeholders (Patton 1997).

This view of the relationship between evaluation and power is derived from the classic organizational theories of Michael Crozier (1964) and James Thompson (1967). Crozier (1964) studied and compared a French clerical agency and tobacco factory. He found that power relationships developed around uncertainties that resulted from information hoarding. Crozier found that supervisors in the clerical agency had no interest in passing information on to their superiors, the section chiefs. Section chiefs, in turn, competed with one another for attention from their superior—the division head. Section chiefs distorted the information they passed up to the division head to enhance their own positions. Section chiefs could get away with distortions because

the lower-level supervisors, who knew the truth, were interested in keeping what they knew to themselves. The division head, on the other hand, used the information he received to schedule production and assign work. Knowing that he was dependent on information from others, and not being able to fully trust that information, his decisions were carefully conservative in the sense that he aimed only at safe, minimal levels of achievement because he knew he lacked sufficient information to narrow risks.

The power of prediction stems to a major extent from the way information is distributed. The whole system of roles is so arranged that people are given information, the possibility of prediction and therefore control, precisely because of their position within the hierarchical pattern. (P. 158)

Whereas Crozier's (1964) analysis centered on power relationships and uncertainties between individuals and among groups within organizations, James Thompson (1967) found that a similar set of concepts could be applied to understand relationships between whole organizations. He argued that organizations are open systems that need resources and materials from outside and that "with this conception the central problem for complex organizations is one of coping with uncertainty" (p. 13). He found that assessment and evaluation are used by people in organizations as mechanisms for reducing uncertainty, enhancing predictability, and increasing their control over the multitude of contingencies with which they are faced. Information for prediction is information for control, thus the power of evaluation. To be power laden, information must be relevant and in a form that is understandable to users. Crozier (1964) recognized this qualifier in linking power to reduced uncertainty: "One should be precise and specify *relevant* uncertainty.... People and organizations will care only about what they can recognize as affecting them and, in turn, what is possibly within their control" (p. 158).

Politics and the Personal Factor

The dominant Weberian perspective in organizational sociology posits that organizations are made up of and operate based on positions, roles, and norms such that the individuality of people matters little because individuals are socialized to occupy

specific roles and positions, and behave according to specific learned norms, all for the greater good of the organization's goal attainment. From this perspective, organizations are an impersonal collection of hierarchical positions. However, people in organizations use evaluation findings. The import of this distinction is illustrated in the findings of a classic study of 20 federal health evaluations that assessed how the findings had been used and sought to identify the factors that affected varying degrees of use (Patton et al. 1977). Respondents were asked to comment on how, if at all, each of 11 factors extracted from the literature on diffusion of innovations and evaluation use had affected use of their study. These factors were methodological quality, methodological appropriateness, timeliness, lateness of report, positive or negative findings, surprise of findings, central or peripheral program objectives evaluated, presence or absence of related studies, political factors, decision maker/evaluator interactions, and resources available for the study. Finally, respondents were asked to "pick out the single factor you feel had the greatest effect on how this study was used."

From this long list of questions only two factors emerged as consistently important in explaining use: (1) political considerations and (2) a factor called "the personal factor." The personal factor is the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates. Where such a person or group was present, evaluations were used; where the personal factor was absent, there was a correspondingly marked absence of evaluation impact.

The personal factor represents the leadership, interest, enthusiasm, determination, commitment, assertiveness, and caring of specific, individual people. These are people who actively seek information to make judgments and reduce decision uncertainties. They want to increase their ability to predict the outcomes of programmatic activity and thereby enhance their own discretion as decision makers, policy makers, consumers, program participants, funders, or whatever roles they play. These are the primary users of evaluation. Studies that were not used stood out in that there was often a clear absence of the personal factor. Thus, the challenge of increasing use has come to consist of two parts: (1) finding and involving those who are, by inclination, information users and (2) training those not so inclined.

Goals-Based Evaluation Research

When Alice encounters the Cheshire cat in Wonderland (Carroll 2006) she asks, [p. II-365 ↓] “Would you tell me, please, which way I ought to walk from here?”

“That depends a good deal on where you want to get to,” said the cat.

“I don't much care where—” said Alice.

“Then it doesn't matter which way you walk,” said the cat.

“—so long as I get *somewhere*,” Alice added as an explanation.

“Oh, you're sure to do that,” said the cat, “if you only walk long enough.” (P. 40).

This story carries a classic evaluation message: To evaluate how well you're doing, you must have some place you're trying to get to. For programs, this has meant having goals and evaluating goal attainment. For evaluators, this means clarifying the intended uses of a particular evaluation. Goals-based evaluation focuses on assessing the extent to which a program attains its stated goals. For rigorous evaluation, program goals should be clear, specific, and measurable. Often, these conditions are not met. Evaluators routinely experience difficulties in assessing goal attainment because of vague and fuzzy goals, conflicts over goals among various stakeholders, and multiple goals articulated without prioritizing. Moreover, distortions can result when program staff pays too much attention to whatever an evaluator decides to measure, essentially giving the evaluator the power to determine what activities become primary in a program. This is expressed in the commonly heard mantra: What gets measured gets done. An example is when teachers focus on having students pass a reading test rather than whether they learn to read. The result can be students who pass mandated competency tests but are still functionally illiterate.

A particularly sociological critique of goals is that they are social constructions that are easily and often reified, that is, they are inherently organizational abstractions treated as if they are real. In an organizational sociology classic, Cyert and March (1963:28)

asserted that *individual people have goals, collectivities of people do not*. They likewise asserted that only individuals can act; organizations or programs, as such, cannot be said to take action. The future state desired by an organization (its goals) is nothing but a function of individual aspirations. This is in keeping with the emphasis on the importance of the personal factor, discussed above, that has taken on prominence in the utilization literature within evaluation research. Individuals use evaluations, not organizations.

Still, organizational sociologists and evaluation researchers find it useful to assume that organizations are purposive despite the difficulties of actually measuring the goals of an organization—that is, treating the organization rather than its individuals as the unit of analysis. Aggregating survey responses from members of an organization doesn't quite make the organization the unit of analysis. Thus, organizational sociologists and evaluation researchers find the purposive image helpful but still elusive. In the end, most evaluation researchers today continue to follow the pragmatic logic of organizational sociologist Charles Perrow (1970) articulated decades ago:

For our purposes we shall use the concept of an organizational goal as if there were no question concerning its legitimacy, even though we recognize that there are legitimate objections to doing so. Our present state of conceptual development, linguistic practices, and ontology offers us no alternative. (P. 134)

Evaluators, like Perrow, are likely to come down on the side of practicality. The language of goals will continue to dominate evaluation. However, the sociological debate clarifies that difficulties in clarifying a program's goals may be due to problems inherent in the notion of goals rather than staff incompetence, intransigence, or opposition to evaluation.

Because of the importance of goals to evaluation research, an evaluation process may begin with an *evaluability assessment* to determine the program's readiness for evaluation. The evaluator works with program managers to help them get ready for evaluation by clarifying goals, finding out various stakeholders' views of important issues, and specifying the model or intervention to be assessed. To do a rigorous and

meaningful evaluation, the evaluator may have to make up for deficiencies in program design. Thus, by default, the evaluator becomes a program or organizational developer.

Evaluators may also be called on to move the unit of analysis from the program to the entire organization. *Mission-oriented evaluation* is an organizational development approach that involves assessing the extent to which the various units and activities of the organization are consistent with its mission, and then determining the degree of mission attainment. In recent years, with an emphasis on creating “learning organizations,” evaluators have been paying increasing attention to the organizational context within which evaluations occur as well as evaluating overall organizational effectiveness and mission attainment.

Turbulent Environments and Goals

How much to seek clarity about goals will depend, among other things, on the program's developmental status and environment. Organizational sociologists have discovered that the clarity and stability of goals are contingent on the organization's environment, especially varying degrees of uncertainty facing the organization. Uncertainty includes things like funding stability, changes in rules and regulations, mobility and transience of clients and suppliers, and political, economic, or social turbulence. What is important about this work from an evaluation perspective is the finding that the degree of uncertainty facing an organization directly affects the degree to which goals and strategies for attaining goals can be made concrete and stable. The less certain the environment, the less stable and less concrete the organization's goals will be. Effective [p. II-366 ↓] organizations in turbulent environments adapt their goals to changing demands and conditions.

Various Evaluation Research Purposes

Evaluation findings can serve three primary purposes: rendering judgments, facilitating improvements, and/or generating knowledge. Judgments are undergirded by the accountability perspective; improvements are informed by a developmental perspective; and generating knowledge operates from the knowledge perspective of academic

values. These are by no means inherently conflicting purposes and some evaluations strive to incorporate all three approaches, but one of these purposes is likely to become the dominant motif in any given effort and prevail as the *primary* purpose informing design decisions and priority uses; or else, different aspects of an evaluation are designed, compartmentalized, and sequenced to address these contrasting purposes. Confusion among these quite different purposes, or failure to prioritize them, is often the source of problems and misunderstandings in evaluation, and can become disastrous at the end when it turns out that different intended users had different expectations and priorities.

In judgment-oriented evaluations, specifying the criteria for judgment is central and critical. Different stakeholders will bring different criteria to the task of judging a program's effectiveness. *Summative evaluation* constitutes an important purpose distinction in any menu of alternative evaluation purposes. Summative evaluations judge the overall effectiveness of a program and deal with the problem of attributing measured results to the program intervention. Summative evaluations are particularly important in making decisions about continuing or terminating an experimental program or demonstration project. As such, summative evaluations are often requested by funders. In judgment-oriented evaluations, the logic of valuing rules. Four steps are necessary: (1) select criteria of merit; (2) set standards of performance; (3) measure performance; and (4) synthesize results into a judgment of value. This is clearly a deductive approach.

Summative evaluation contrasts with *formative evaluation*, which focuses on ways of improving and enhancing programs rather than rendering definitive judgment about effectiveness. In contrast to summative evaluations, improvement-oriented (formative) evaluations often use an inductive approach in which criteria are less formal as one searches openly for whatever areas of strengths or weaknesses may emerge from looking at what's happening in the program.

Using evaluation results to improve a program turns out, in practice, to be fundamentally different from rendering judgment about overall effectiveness, merit or worth. Improvement-oriented forms of evaluation include formative evaluation, quality enhancement, responsive evaluation, learning organization approaches, humanistic evaluation, and total quality management, among others. What these approaches

share is a focus on making things better rather than rendering summative judgment. Judgment-oriented evaluation requires explicit criteria and values that form the basis for judgment. Improvement-oriented approaches tend to be more open-ended, gathering varieties of data about strengths and weaknesses with the expectation that both will be found and each can be used to inform an ongoing cycle of reflection and innovation. Program management, staff, and sometimes participants tend to be the primary users of improvement-oriented findings, whereas funders and external decision makers tend to use summative evaluation. Improvement-oriented evaluations aim to determine the program's strengths and weaknesses, the extent to which participants are progressing toward desired outcomes, which types of participants are making good progress and which types aren't doing so well, and what kinds of implementation problems have emerged. The formative evaluator looks for unexpected consequences and possible side effects. It is especially important to gather data about how staff and clients are interacting, and to gather data on staff and participant perceptions of the program, finding out what they like, dislike, and want to change. Data on perceptions of the program's culture and climate may be part of the evaluation. The evaluation may examine how funds are being used compared with initial plans and how the program's external environment is affecting internal operations, looking for efficiencies that might be realized. In formative evaluation, it is especially important to gather evaluative feedback from program participants who receive services and to take that feedback seriously.

One classic metaphor explaining the difference between summative and formative evaluation is that when the cook tastes the soup, that's formative; when the guests taste the soup, that's summative.

Both summative and formative evaluations involve the *instrumental* use of results. Instrumental use occurs when a decision or action follows, at least in part, from the evaluation. Evaluations are seldom the sole basis for subsequent summative decisions or program improvements, but, when well done, they can contribute, often substantially, to programmatic decision making.

Alternative Ways of Focusing Evaluations

As just noted, different types of evaluation research can ask different questions and focus on different purposes. Various options can be and often are used together within the same evaluation, or options can be implemented in sequence over a period of time, for example, doing implementation evaluation before doing outcomes evaluation, or formative evaluation before summative evaluation. Below are some examples of alternative evaluation types (left column) and their defining question in relation to key sociological issues (right column, italics). [p. II-367 ↓]

| Focus or Type of Evaluation | Defining Question/Approach and Sociological Contribution |
|---|--|
| Accreditation focus | Does the program meet minimum standards for accreditation or licensing? |
| | Sociology of professions |
| Compliance focus | Are rules and regulations being followed? |
| | Sociological sensitivity to formal rules versus informal norms |
| Comparative focus and contextual analysis | How do two or more programs compare in different contexts? What are the environments within which the program operates politically, socially, economically culturally, and scientifically? How does this context affect program effectiveness? |
| | Comparative sociology |
| Diversity focus | The evaluation gives voice to different perspectives on and illuminates various experiences with the program. |
| | Constructivist sociology and sociology of knowledge |

| | |
|---------------------------------|--|
| Empowerment evaluation | The evaluation is conducted in a way that affirms participants' self-determination and builds capacity to evaluate themselves. |
| | Sociological sensitivity to power differentials |
| Longitudinal evaluation | What happens to the program and to participants over time? |
| | Methodological issues in longitudinal sociological designs |
| Norm-referenced approach | How does this program population compare to some specific norm or reference group selected variables? |
| | Comparative sociology |
| Social and community indicators | What routine social and economic data should be monitored to assess the impacts of this program? What is the connection between program outcomes and larger-scale social indicators, for example, crime rates? |
| | Community sociology |
| Utilization-focused evaluation | What information is needed and wanted by primary intended users that will actually be used for program improvement and decision making? |
| | Sociology of knowledge |

Conceptual Use of Evaluation Findings

The preceding examples offer various ways of focusing evaluations to achieve what earlier was emphasized as the *instrumental* use of findings, that is, using evaluations to

make program improvements and overall summative judgments about a program's merit or worth. *Conceptual use of findings*, on the other hand, contrasts with instrumental use in that no decision or action is expected; rather, it is the use of evaluations to influence thinking about issues in a general way. The evaluation findings contribute by increasing knowledge. This knowledge can be as specific as clarifying a program's model, testing theory, distinguishing types of interventions, figuring out how to measure outcomes, generating lessons learned, and/or elaborating policy options. One form of conceptual use is called "enlightenment," a distinction aimed at describing the effects of evaluation findings being disseminated to the larger policy community where they may affect the terms of debate. Generalizations from evaluation research can become part of the knowledge base for policy making. Case studies of evaluations and decisions tend to show that generalizations and ideas that come from research and evaluation help shape the development of policy.

One formal knowledge-oriented approach is called *theory-driven evaluation*. This connection of evaluation research to social science theory tends to focus on increasing knowledge about how effective programs work in general. For example, results from evaluations can contribute to theories about how to solve societal problems or produce important sustainable social innovation. Theorydriven evaluation can be aimed at particular aspects of the programming process, for example, implementation theory aimed at better understanding the nature of program delivery. Such knowledge-generating efforts focus beyond the effectiveness of a particular program to future program designs and policy formulation in general.

As the field of evaluation has matured and a vast number of evaluations has accumulated, the opportunity has arisen to look across findings about specific programs to formulate generalizations about effectiveness. This involves synthesizing findings from different studies. An early and important example of synthesis evaluation was Lisbeth Schorr's (1988:256–83) *Within Our Reach*, a study of programs aimed at breaking the cycle of poverty. She identified "the lessons of successful programs" as follows:

- Offering a broad spectrum of services
- Regularly crossing traditional, professional, and bureaucratic boundaries

- Seeing the child in the context of family and the family in the context of its surroundings, that is, holistic approaches
- Coherent and easy-to-use services
- Committed, caring, results-oriented staff
- Finding ways to adapt or circumvent traditional professional and bureaucratic limitations to meet client needs
- Professionals redefining their roles to respond to severe needs
- Overall, intensive, comprehensive, responsive, and flexible programming

Such generalizable evaluation findings about principles of effective programming have become the knowledge base of the field of evaluation research. Being knowledgeable about patterns of program effectiveness allows evaluators to provide guidance about development of new initiatives, policies, and strategies for implementation. These kinds of “lessons” constitute accumulated wisdom—principles of effectiveness or “best practices”—that can be adapted, indeed must be adapted, to specific programs, or even entire organizations.

In this vein, a special evaluation issue of *Marriage and Family Review* was devoted to “Exemplary Social Intervention Programs” (Guttman and Sussman 1995) not only looking at specific examples but also extracting cross-case patterns and principles. Such qualitative syntheses in evaluation have become increasingly important as policymakers look beyond the effectiveness of specific programs to more generic principles of effectiveness based on “high-quality lessons learned” (Patton 2002:564–566).

Sociology and Evaluation Research

Sociology has contributed to evaluation research methodologically, through theory construction, and substantively, by informing critical questions and deepening evaluative inquiry.

Sociological areas of specialization that have made important contributions to evaluation research include the sociology of knowledge; organizational sociology, conflict theory; and areas related to special efforts at societal intervention that are the

object of programming and therefore evaluation research, for example, criminology, gerontology, marriage and family studies, sociology of youth, and community sociology. Sociologists like Edward Suchman (1967), Carol Weiss (1972, 1977), Michael Q. Patton (1978), Peter Rossi and Howard Freeman (1982) helped create the interdisciplinary field of evaluation research. Evaluation research can be viewed as a particular and specialized arena within applied sociology. As evaluation research has grown and matured into a recognized profession, it has also matured into an important arena of sociological practice.

In the future, evaluation research will be aiming to increase use beyond projects and programs as primary units of analysis to evaluating overall organizational effectiveness and the impacts of social policies, thereby having greater influence on policy (Rossi, Lipsey, and Freeman 2004; Weiss, Murphy-Graham, and Birkeland 2005) and a broader range of audiences (Baxter and Braverman 2004). The cross-cultural and global reach of evaluation will accelerate with more attention to “contextually responsive evaluation frameworks” (Thomas and Stevens 2004), training evaluators to work in culturally diverse settings (Thompson-Robinson, Hopson, and SenGupta 2004) and adapting evaluation practices and standards to international settings (Russon and Russon 2004). Attendant to these developments will be increased emphasis on getting feedback from program participants about the services they receive and using participatory evaluation processes in which both program staff and intended beneficiaries play a meaningful role in the evaluation process (Fetterman and Wandersman 2005). The emergence of evaluation research as an identifiable field of professional practice and scholarship will be solidified as evaluation knowledge is codified and disseminated (Alkin 2004; Mathison 2005) and essential evaluator competencies are crystallized (Stevahn et al. 2005). Technology and global communications will also surely influence the future of evaluation research. To keep up with these developments, some key Web sites for tracking future developments in this still emergent field of evaluation research are offered in [Table 41.1](#).

| | |
|--|---|
| American Evaluation Association | http://www.eval.org |
| Journal of Multidisciplinary Evaluation (an online-only journal) | http://evaluation.wmich.edu/jmde |

| | |
|---|---|
| Evaluation Checklists | http://www.wmich.edu/evalctr/checklists |
| International Organization for Cooperation in Evaluation | http://www.internationalevaluation.com |
| International Development Evaluation Association | http://www.ideas-int.org |
| Archives of EvalTalk, the American Evaluation Association Discussion List (LISTSERV 14.5) | http://bama.ua.edu/archives/evaltalk.html |

MICHAEL QUINNPATTON *Utilization-Focused Evaluation*, Saint Paul, Minnesota

Further Reading

Alkin, ed. , Marvin, ed. 2004. *Evaluation Roots: Tracing Theorists' Views and Influences* . Thousand Oaks, CA: Sage.

Baxter, L. W., and M. T.Braverman. 2004. "Communicating Results to Different Audiences." Pp. 281–304 in *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy* , edited by M. T. Braverman, N. A. Constantine, and J. K. Slater. San Francisco, CA: Jossey-Bass.

Campbell, Donald T. "Methods for the Experimenting Society." *Evaluation Practice* 12 (3) 223–60 1991. <http://dx.doi.org/10.1016/0886-1633%2891%2990039-Z>

Carroll, Lewis. 2006. *Alice in Wonderland* . Ann Arbor, MI: Ann Arbor Media.

Crozier, Michel. 1964. *The Bureaucratic Phenomenon* . Chicago, IL: University of Chicago Press.

Cyert, Richard, and James G.March. 1963. *A Behavioral Theory of the Firm* . Englewood Cliffs, NJ: Prentice Hall.

Fetterman, ed. , David and Abraham Wandersman, eds. 2005. *Empowerment Evaluation Principles in Practice* . New York: Guilford Press.

Guttman, David, ed. , and Marvin B. Sussman, eds. 1995. "Exemplary Social Intervention Programs for Members and Their Families" (Special issue). *Marriage and Family Review* 21(1,2). http://dx.doi.org/10.1300/J002v21n01_01

Mathison, ed. , Sandra, ed. 2005. *Encyclopedia of Evaluation* . Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412950558>

Mertens, D. "The Inauguration of the International Organization for Cooperation in Evaluation." *American Journal of Evaluation* 26 (1) 124–30 2005. <http://dx.doi.org/10.1177/1098214004273186>

Mills, C. Wright. 1959. *The Sociological Imagination* . New York: Oxford University Press.

Patton, Michael Q. 1978. *Utilization-Focused Evaluation* . Beverly Hills, CA: Sage.

Patton, Michael Q. 1997. *Utilization-Focused Evaluation: The New Century Text* . 3d ed. Thousand Oaks, CA: Sage.

Patton, Michael Q. 2002. *Qualitative Research and Evaluation Methods* . 3d ed. Thousand Oaks, CA: Sage.

Patton, Michael Q., Patricia S.Grimes, Kathryn M.Guthrie, Nancy J.Brennan, Barbara D.French, and Dale A.Blyth. 1977. "In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research." Pp. 141–64 in *Using Social Research in Public Policy Making* , edited by C. Weiss. Lexington, MA: D. C. Heath.

Perrow, Charles. 1970. *Organizational Analysis: A Sociological View* . Belmont, CA: Wadsworth.

Rogers, Everett. 1962. *Diffusion of Innovation* . New York: Free Press.

Rogers, Everett M., and LynneSvenning. 1969. *Managing Change* . San Mateo, CA: Operation PEP.

Rogers, Everett M., and Floyd F. Shoemaker. 1971. *Communication of Innovation*. New York: Free Press.

Rossi, Peter H., and H. E. Freeman. 1982. *Evaluation: A Systematic Approach*. Beverly Hills, CA: Sage.

Rossi, Peter H., M. Lipsey, and H. E. Freeman. 2004. *Evaluation: A Systematic Approach*. 7th ed. Thousand Oaks, CA: Sage.

Russon, C., ed., and G. Russon, eds. 2004. *International Perspectives on Evaluation Standards*. *New Directions for Evaluation*, No. 104. San Francisco, CA: Jossey-Bass.

Schorr, Lisbeth. 1988. *Within Our Reach: Breaking the Cycle of Disadvantage*. New York: Doubleday.

Stevahn, L., J. King, G. Ghere, and J. Mnnema. "Establishing Essential Competencies for Program Evaluators." *American Journal of Evaluation* 26 (1) 43–59 2005. <http://dx.doi.org/10.1177/1098214004273180>

Stufflebeam, Daniel. "An Interview with Daniel L. Stufflebeam." *Educational Evaluation and Policy Analysis* 2 4 1980.

Suchman, Edward A. 1967. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York: Russell Sage.

Thomas, V. G., ed., and F. I. Stevens, eds. 2004. *Co-Constructing a Contextually Responsive Evaluation Framework*. *New Directions for Evaluation*, No. 101. San Francisco, CA: Jossey-Bass.

Thompson, James D. 1967. *Organizations in Action*. New York: McGraw-Hill.

Thompson-Robinson, M., ed., R. Hopson, ed., and S. SenGupta, eds. 2004. *In Search of Cultural Competence in Evaluation*. *New Directions for Evaluation*, No. 102. San Francisco, CA: Jossey-Bass.

Weiss, ed. , Carol, ed. 1977. Using Social Research in Public Policy Making . Lexington, MA: D. C. Heath.

Weiss, Carol H. 1972. Evaluation Research: Methods of Assessing Program Effectiveness . Englewood Cliffs, NJ: Prentice Hall.

Weiss, C., E. Murphy-Graham, and S.Birkeland. "An Alternative Route to Policy Influence." American Journal of Evaluation 26 (1) 12–30 2005. <http://dx.doi.org/10.1177/1098214004273337>

<http://dx.doi.org/10.4135/9781412939645.n98>