

Review of Educational Research

<http://rer.aera.net>

Issues in the Evaluation of Social Programs

Francis G. Caro

REVIEW OF EDUCATIONAL RESEARCH 1971 41: 87

DOI: 10.3102/00346543041002087

The online version of this article can be found at:

<http://rer.sagepub.com/content/41/2/87>

Published on behalf of



American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *Review of Educational Research* can be found at:

Email Alerts: <http://rer.aera.net/alerts>

Subscriptions: <http://rer.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

Citations: <http://rer.sagepub.com/content/41/2/87.refs.html>

ISSUES IN THE EVALUATION OF SOCIAL PROGRAMS

Francis G. Caro

*University of Colorado*¹

Efforts in the past decade to reduce the incidence of major social problems have stimulated a new interest in program evaluation. Numerous writings have appeared on the use of the concepts and methods of behavioral research in evaluating these interventions. This paper is a review of that literature. It surveys evaluation programs designed to produce some identifiable change in individual or social groups. The first part of the paper is concerned with basic issues which include definitions, approaches to evaluation methodology, roles of evaluation in program development, and distinctions among various forms of research. The second section deals with organizational matters such as the establishment of the evaluative research role, administration of evaluative research, utilization of the results of evaluation, and implications of client activism for evaluation. The third portion reviews methodological issues in measurement and design of evaluation studies.

Basic Considerations

Definitions

Attempts to define evaluation reflect concern with both the results and the desirability or value of programs. Greenberg (1968), Brooks (1965), and Suchman (1967a) emphasized the information-seeking aspect of evaluation. Greenberg (1968, p. 260) defined it as "the procedure by which programs are studied to ascertain their effectiveness in the fulfillment of goals." Brooks (1965, p. 34) listed as evaluation objectives the determination of: (a) the extent to which the program achieves its goal, (b) the relative impact of key program variables, and (c) the role of the program as contrasted to external variables. Suchman (1967a, pp. 31-32) defined evaluation as "the determination . . . of the results . . . attained by some activity . . . designed to accomplish some valued goal or objective." He also identified (1966, p. 68) four evaluation categories: (a) effort (the amount of action), (b) effect (results of effort), (c) process (how an effect was achieved), and (d) efficiency (effects in relation to cost).

¹The author is now at Brandeis University.

The judgmental dimension was emphasized by Scriven (1967, pp. 40-41) who defined evaluation as a "methodological activity which combines performance data with a goal scale." Glass (1971) similarly stressed that evaluation is an attempt to assess the worth or social utility of a thing. He argued that since the desirability of announced program goals might be questioned, evaluation should include procedures for the evaluation of those goals.

Methodological Approaches

Several distinctly different approaches to evaluation methodology can be identified. Legislators, administrators, practitioners, recipients of services, and journalists are among those who typically rely on impressionistic or informal evaluation. Stake (1967, pp. 23-24) described informal evaluation as dependent on casual observation, implicit goals, intuitive norms, and subjective judgment; he characterized it as of variable quality—sometimes penetrating and insightful, sometimes superficial and distorted. Similarly, Mann (1969) noted that observations by participants may provide suggestive leads for interpreting the effects of programs, but because the extent of their bias is unknown, it is impossible to judge the accuracy of their conclusions.

Among formal approaches to evaluation, a distinction can be made between those emphasizing inputs and those emphasizing outputs. Educational accrediting agencies, municipal building inspectors, and fire insurance underwriters base their evaluative judgments on inputs, using explicit checklists and formulas. Educational programs are evaluated on the basis of such factors as teacher qualifications and ratio of library books to students; adequacy of housing is judged on the basis of plumbing facilities and sleeping arrangements; fire insurance raters consider the nature of building materials and fire fighting equipment. Data are typically obtained through site inspections. Glass (1971) pointed out that since it is based on subjective judgments, this approach is weak in the areas of objectivity and validity.

The program accounting approach to evaluation also emphasizes input of efforts. It focuses on the maintenance and quantitative analysis of project activity records. The extent of actual practitioner-client contact and the number of clients exposed to programs are typical concerns. Outputs or effects tend to receive little attention; program accounting is tied to routine agency records and agencies can rarely undertake the extensive follow-up activities necessary for complete information on the outcome of services. However, since it can provide information on the ability to establish contacts with clients and the cost of program-client contacts, program accounting is useful as a procedure for determining the administrative viability of programs.

An emphasis on outputs or effects and a concern with the use of the scientific method characterizes evaluative research, the approach of primary concern here. Suchman (1969, p. 15) distinguished between evaluation, a "general social process of making judgments of worth regardless of the basis for such judgments," and evaluative research, the "use of the scientific method for collecting data concerning the degree to which some specified activity achieves some desired effect." Similarly, Hyman and Wright (1967, p. 742) called for evaluation based on "methods that yield evidence that is objective, systematic, and comprehensive." The emphasis of evaluative research on outputs need not imply a lack of concern for input variables. Scriven (1967), for example, suggested "mediated" evaluation as a way of combining input and output variables to permit study of the process through which goals are pursued.

Program Development and Administration

Evaluation may be viewed as a phase in a systematic program development. Ideally, action programming is preceded by a planning process that includes: (a) identification of problems, (b) specification of objectives, (c) analysis of the causes of problems and the shortcomings of existing programs, and (d) an examination of possible action alternatives. Evaluation follows program implementation and provides a basis for further planning and program refinement. (Although evaluation may follow implementation, it is, of course, desirable that evaluation activities begin prior to implementation.) The planning-action-evaluation cycle may be repeated indefinitely until objectives are realized or problems and objectives are redefined. Results of evaluation may be used to modify programs already in progress to increase the likelihood of realization of long-term goals. When evaluation is viewed as part of a process of planned change, the utilization of evaluation findings in decision making becomes a key concern.

Evaluative research is concerned with stable and well established programs as well as with new programs for which viable administrative patterns are being sought. Scriven (1967, p. 43) introduced the terms "formative" and "summative" to distinguish between these two evaluation concerns. Formative evaluation is designed to improve a program while it is still fluid; summative evaluation is designed to appraise a product after it is well established.

A predisposition toward gradual and moderate change is thought to produce an interest in evaluative research. Where change is thought to be undesirable or impossible, little interest in evaluation can be expected from the guardians of societal institutions. Groups demanding rapid and radical change are also unlikely consumers of evaluative research; their inclinations are likely to be ideological rather than empirical and evaluative

researchers are not likely to respond to their information needs rapidly enough.

Emphasis on evaluative research is most appropriate when program effects cannot be expected to be directly and immediately evident (Coleman, 1969). Such is often the case in contemporary large-scale education, welfare, and social service programs whose effects are often subtle and diffuse. Large-scale programs also tend to increase the physical and social distance between policy makers and recipients of services. When they are in close contact with clients, policy makers may have reason to be confident of their own informal evaluations of programs. As their contacts with the client population decrease, however, policy makers may recognize the need for more formal evaluation procedures (Trow, 1967).

In principle, evaluation activities may generate judgments regarding effectiveness of performance on such varied levels as programming approaches (e.g., remedial reading or income-maintenance programs), administrative units (eg., schools, departments, or agencies), individual practitioners (e.g., physicians or teachers), and recipients of services (e.g., patients, clients, or students) (Cronback, 1963). In practice, for reasons discussed below, evaluators whose primary concern is with program effectiveness usually deliberately avoid making judgmental statements regarding particular administrative units, practitioners, or recipients of services.

Evaluation may be considered a programming input, subject to evaluation in the same manner as other inputs. In cost-benefit terms, the cost of evaluation should be related to the benefits which evaluation data and judgments contribute to programming efficiency or effectiveness. A heavy investment in formal evaluation is most likely to be justified when a program is expensive, its impact is potentially great but uncertain, and the potential for diffusion of programming concepts is great. Glass (1971) contended that decisions to conduct evaluative research should reflect estimates of the cost of evaluation, the extent to which there is uncertainty regarding program effectiveness, and the cost of implementing alternate programs.

Theoretically, evaluative research may be undertaken without any formal sponsorship, may be based on a wide range of value perspectives, and its findings may be reported to a variety of audiences. In practice, however, because of problems of cost and access to information, formal evaluation is usually a sponsored activity. External funding agencies (e.g., private foundations or the federal government) and top administrators of action organizations are the most common sponsors. Whether conducted by an internal unit of an organization or by outside consultants, evaluative researchers usually are directly linked to persons high in the administrative structure of the action organization. Those who actually carry out the pro-

grams to be evaluated are, then, subordinate to those to whom evaluators report. The issues addressed by evaluation and the manner in which results are reported are strongly related to sponsorship. Consequently, the interests of the general public, practitioners, and recipients of services are often not fully served by evaluators.

Science, Research, and Evaluation

From a behavioral science point of view, evaluative research represents an application of the scientific method which is quite different from that of basic research. Some insist on a sharp distinction between research and evaluation, while others classify evaluation as a form of research. Wrightstone (1969, p. 5) suggested that "Research is more concerned with the basic theory and design of a program over an appropriate period of time, with flexible deadlines, and with sophisticated treatment of data that have been carefully obtained." Evaluation, on the other hand, "may be concerned with basic theory and design, but its primary function . . . is to appraise comprehensively a practical . . . activity to meet a deadline" Suchman (1969) argued that the distinction between basic research and evaluative research is one of purpose rather than method. Evaluative research applies the scientific method to problems which have administrative consequences while basic research is concerned with problems of theoretical significance. Cherns (1969) distinguished between pure basic research, which is stimulated by the perceived needs of an academic discipline, and action (evaluative) research, which is concerned with an ongoing problem in an organizational framework and involves the introduction and observation of planned change. Cherns also pointed to differences in diffusion and generality between the types of research. In the case of basic research, the potential for generality is great but the potential for immediate utilization is limited. Evaluative research has a limited potential for generality but the potential for immediate utilization is great.

Evaluative research represents only one form of applied or action research; research may contribute to social action without assessing the effects of specific interventions. Research on the causes of problem behavior, on the incidence and concentration of patterns of social problems, and on public knowledge of and attitudes toward existing services may all have important policy implications without being specifically evaluative.

For social scientists interested in contributing to programs of direct change, evaluative research is only one possible role. Alternately, social scientists may contribute to training programs and engage in consulting activities. Brooks (1965) suggested that social scientists provide ideas for experimentation and encourage the greatest possible rationality in the planning process. They can aid in the identification of objectives and action

alternatives and the prediction of consequences of possible courses of action. (See Bennis, 1965; Likert & Lippitt, 1953.)

The Organizational Context

Problems in Establishing and Maintaining the Evaluative Research Role

Although the rationale for a central role for formal behavioral evaluation is often strong, effective participation by evaluative researchers in social programming is much less common. When evaluation is examined from an organizational and occupational perspective, some of the practical problems in establishing and maintaining the evaluation role become apparent.

Of fundamental importance is the fact that, traditionally, decision makers have not accorded evaluative research a major role in policy formation and change in social programming (Rossi, 1969). Policy has been formed without considering the kinds of evaluation data needed to sustain the worth of a program. Objective evidence of the effects of programs has not been demanded as a basis for modifying programs. Satisfied with informal evaluation, administrators have often included evaluative research only when it was required by a funding agent. Recent emphasis on evaluation in education, for example, stems largely from a provision of the Elementary and Secondary Education Act of 1965.

Some administrators may consider evaluative research expensive and of little practical value. They may also have important covert reasons for resisting formal evaluation. For example, the presence of an evaluation component invites administrators to consider the possibility that their policies do not lead to the effective realization of announced objectives. Because administrative claims for programs are usually extremely optimistic, evaluative research results almost inevitably are disappointing (Rossi, 1967). Campbell (1969) observed that ambiguous results help to protect administrators where there is a possibility of failure. Freely available facts might reduce the privacy and security of some administrators, making them vulnerable to inquiries about their honesty and efficiency. In addition, administrators may resent evaluators who raise questions about basic organizational premises or suggest evaluative criteria which may be embarrassing to the organization.

Horowitz (1969) identified several other reasons why administrators may consider social scientists who belong to an internal research unit troublesome members of an action organization. Social scientists often demand preferential treatment which is resented by other employees. Presuming superior wisdom on their own part, social scientists may seek direct access to top decision makers, thereby threatening bureaucrats who are

by-passed. Further, the extracurricular involvement of social scientists in writing, teaching, and the like are often resented.

At the same time, administrators interested in evaluative research have often found it difficult to recruit and hold qualified behavioral scientists. Like other scientists, behavioral scientists often prefer to be oriented toward the general scientific community rather than the needs and goals of the organization which employs them (McKelvey, 1969). Scientists typically wish to do research that contributes to a scientific body of knowledge. Administrators, however, typically expect scientists whom they employ to do research which contributes directly to the goals of their organizations. Social scientists who may wish to publicize their work also resent the norm of secrecy prevailing in some organizations (Horowitz, 1969). In contrast to administrators who want social scientists to work within the framework of established policy, social scientists may want to challenge an agency's ideological premises (Horowitz, 1969). In addition, some social scientists have been concerned that agreement to undertake the evaluation of a program may be interpreted as implicit commitment to the philosophy or goals of that program; they may give a program a legitimacy they do not believe it deserves (Ferman, 1969). Other deterrents have been the low prestige accorded to applied research in academic settings, exasperation with the methodological and administrative problems of conducting research in an action setting, and disagreements regarding the use of research results.

Problems in Administration of Evaluative Research

Successful administration of evaluative research depends on cooperation from agency administrators and practitioners charged with implementing action programs. Even though they often advocate extensive collaboration and communication with administrators, evaluative researchers typically insist that ultimate responsibility for research design and execution is theirs. Administrative interference with what social scientists consider critical issues in the design and execution of research is seriously resented (Smith, Sim, & Bealer, 1960).

Whether or not researchers are agency employees, they are readily drawn into staff-management conflicts. A number of observers noted that acceptance of evaluative research at upper management levels was often accompanied by suspicion of research at lower levels (Rodman & Kolodny, 1964; Lippitt, Watson, & Westley, 1958; Argyris, 1958; Whyte & Hamilton, 1964; Likert & Lippitt, 1953). Because evaluation is linked to top administration and involves examination of the activities of staff subordinates, evaluators are sometimes accused of being management spies. Staff practitioners interested in avoiding criticism of their work are likely to attempt to conceal real or imagined shortcomings. Such steps, of course, would add greatly to the evaluator's difficulties in obtaining valid data.

Research neutrality is also likely to pose a problem with practitioners who consider a strong value commitment to their programs important. Scriven (1967), for example, reported that some practitioners feel that the skepticism of evaluative researchers may dampen the creativity of a productive group. Argyris (1958) argued that research neutrality leads to subject alienation which then produces anxieties in the researcher, resulting in invalid observations.

The purely mechanical demands of data collection may also create a burden. Practitioners, perhaps correctly, consider themselves overburdened with record keeping. Typically, even when record keeping is emphasized, records are not sufficiently accurate or complete to satisfy research criteria. Conflict between research and service goals may also interfere with the collection of research data. When a research design calls for action inconsistent with immediate service goals, practitioners may disregard research needs in favor of providing services. Compounding the problem, practitioners may neglect to inform evaluators that clients were shifted from a control or comparison group to an experimental group.

Different conceptions of efficient use of time may lead to mutual annoyance. A professional evaluator, for example, is not accustomed to turning in daily time sheets, but his failure to do so may be interpreted as a sign of indolence by an administrator with time and cost concerns. Bynder (1966, p. 67), reflecting on his research in a social work unit of a general hospital, observed that "Thinking is not a tangible use of time, and therefore, could not be accepted in an agency which measured work in terms of clients interviewed, physicians contacted, meetings attended, and pages written." An insecure social scientist may respond by engaging in activities which make him appear busy but which are detrimental to long-term evaluation objectives.

Status ambiguities may further strain relationships, especially if an evaluator has had more formal education but less clinical experience than administrator and practitioner counterparts. The social scientist may bring to the situation an academic disrespect for practical problems; administrators and practitioners, in turn, may be defensive about their educational inferiority and highly sensitive to what they interpret as the snobbism of evaluative researchers. Deteriorating relationships have led threatened practitioners to claim that evaluators are incompetent because they do not understand the practical problems of an action agency. The evaluators, perceiving themselves as exposed and defenseless members of a minority group in the action organization, sometimes react at this point by looking for ways to return to an academic setting.

Two basic problems were discussed in the literature concerning the publication of evaluative research results (see, e.g., Rodman & Kolodny, 1964). Agencies often impose controls on the publication of "sensitive"

data because a negative report may threaten not only the agency's public image but also its access to funds. If it is agreed that project results should be reported, there may be disagreements about publication credits. The evaluator researcher, who contributed the research design, data analysis, and write-up, may regard the report as a scientific publication for which he is solely responsible. The administrator, emphasizing the content of the project, may believe he deserves major recognition for conceiving and implementing the program.

A final important issue is the availability of the funds needed for evaluative research. Action organizations nearly always operate on tight budgets. Administrators must attempt to use funds to provide as much service as possible. The cost of the elaborate data collection and analysis essential to evaluate research may represent a substantial proportion of the total project budget. Given the often intangible and uncertain contribution of evaluative research, requests for evaluation funds may be among the first to suffer in times of budget curtailment.

Problems in Utilization of Results of Evaluation

Since the ultimate purpose of evaluation is a contribution to the effectiveness of action programs, implementation of research results is a critical phase in the evaluation process. Yet numerous writers have warned that even the most carefully designed and executed evaluative research does not automatically lead to meaningful action. (For some examples of cases in which findings of evaluative research were ignored or rejected by program administrators, see Rossi (1967) and Hall (1966).) Disregard for results of evaluation appears to stem from a variety of sources.

Some of the nonuse of evaluation results is attributable to limitations of the research itself. In discussing demonstration projects, Rein and Miller (1967) noted that evaluative research often cannot produce results early enough to be a major factor in short-term policy decisions. Mann (1969) similarly reflected on the dilemmas of rigor, timing, and utility of evaluation: "The better the study, the longer it takes, and consequently the less usefulness it may have. Conversely, the sloppier the procedure, the more likely it is to provide information on questions of interest even though this data will be of doubtful validity [p. 13]." Weiss (1966) indicated that the influence potential of evaluation may be limited because results are indefinite, show only small changes, and fail to indicate the relative effectiveness of various components or the reasons for a program's success or failure.

Of basic importance in cases where pertinent evaluation results are ignored is the evaluator's lack of authority. Since the evaluator's capacity is that of an advisor, policy makers are under no obligation to accept his recommendations. Nonuse of evaluation findings is sometimes explained by the fact that evaluation was included for the "wrong reasons." Downs

(1965) pointed out that professional advice is sometimes sought to justify decisions already made or to postpone action. Several commentators suggested that an evaluation component is sometimes supported because it lends an aura of prestige to an action enterprise (Rodman & Kolodny, 1964; Bynder, 1966; Rosenthal & Weiss, 1966; Schulberg & Baker, 1968). An administrator may support an evaluator in the hope that the evaluator will provide other services—for example, the organization of information to justify grant requests (Miller, 1965a; Luchterland, 1967). As previously indicated, evaluation sometimes is included in action programs only because it is required by law or the administrative regulations of a funding agent. In these cases evaluation results may be ignored because administrators do not adequately understand or appreciate their relevance or, perhaps, because they resent evaluation as an imposition. Discrepancies between the findings of evaluative research and informal evaluations, the personal convictions and professional ideologies of decision makers, and judgments of the competence of evaluators also contribute to the nonuse of evaluative research findings (Sadofsky, 1966).

Disagreements regarding evaluative criteria sometimes contribute to nonuse of findings. Rossi (1969) observed that administrators sometimes discount evaluation findings by claiming that the “real” goals of the project were not measured. Schulberg and Baker (1968) questioned the wisdom of the usual practice of building evaluation on the public goals of an organization because administrators may have no intention of achieving those goals. An evaluative researcher, then, may be ineffective because he misread the administrator’s real intent.

Basic Tensions Between Evaluative Researchers and Administrators

It has been shown that strained relations between evaluative researchers and administrators often occur in the introduction, execution, and utilization of evaluative research. Many of the specific obstacles to effective collaboration can be summarized by consideration of several basic orientations in which administrators and evaluative researchers are likely to differ markedly: service versus research, specificity versus generality, methods, status quo versus change, explanations for failures, and academic versus practical experience.

Service versus research. In contrast to the practitioner who is concerned with the immediate and specific application of knowledge, the evaluative researcher is responsible for the acquisition of knowledge. The service-research tension is most evident in field settings where research and service perspectives call for opposite courses of action. An evaluative research design may call for the assignment of a client to a control group when, from a service perspective, it appears preferable that he receive the experimental treatment (Argyris, 1960; Freeman, 1963; Perry & Wynn,

1959). In addition, evaluative researchers, reflecting their academic backgrounds, are likely to have a greater appreciation than practitioners for the acquisition of knowledge for its own sake.

Specificity versus generality. In contrast to administrators who emphasize the solution of immediate problems, researchers are often more interested in long-range problem solving. Similarly, administrators emphasize the uniqueness of their agency and program while researchers prefer to generalize in both time and space. What is of theoretical significance to the scientist may be trivial from a practical viewpoint (see Shepard, 1956; Warren, 1963; Merton, 1957; Rodman & Kolodny, 1964; Cherns, 1969).

Methods. Although administrators and researchers usually agree on the use of rational methods in program development, they often do not mean the same thing by "rational." Evaluative research requires explicit statements of objectives and strategies to which administrators find it difficult or undesirable to commit themselves (Schulberg & Baker, 1968). Administrators may be displeased with evaluative research which, in emphasizing organizational outputs, often tends to neglect administrative activities which are needed to maintain the organization as a viable system (Etzioni, 1960). At another level, the researcher's commitment to scientific decision-making procedures may run counter to the administrator's confidence in intuition. Evaluative researchers have a professional interest in being able to show that the scientific method is superior to conventional wisdom as a basis for decision making (Ferman, 1969).

Status quo versus change. Implicit in the evaluation role are attempts to discover inefficiency and to encourage change. Administrators, however, usually prefer to conceal inefficiency and tend to resist disruptive change. A claim to superior knowledge of human affairs predisposes social scientists to dramatize inadequacies in the conventional wisdom on which programs are often based. Administrators, however, look for evidence of success in past and current programs to assert their competence. Evaluative researchers are, thus, predisposed to see a need for fundamental change while administrators are inclined to defend the programs with which they have been identified (see Argyris, 1958; Ferman, 1969). (When they are new to their positions, however, administrators may be receptive to suggestions from evaluators for basic changes since major weaknesses in programs can be attributed to a previous administration.)

Explanations of failure. Evaluators and administrators frequently emphasize different explanations for the persistence of social problems. Again, apparently because of a desire to assert their competence, administrators tend to accept the validity of the theoretical premises on which their programs are based. Attributing failure to the inadequate resources available for the application of his approach, the administrator is likely to call for an

expansion of present efforts. Evaluators who are free to question program premises often attribute failure to an inadequate understanding of the basic problem. They are likely to suggest that a radically different programming approach is needed in order to address the problem effectively. (In arguing that what is needed is "more of the same," the practitioner may also serve his professional interest in expanding the demand for his services. Evaluative researchers similarly have a vested professional interest when they argue that more effective programming requires an expanded emphasis on evaluation.)

When administrators and evaluators both acknowledge difficulties in implementing programs, administrators are likely to look for explanations which are individual (e.g., incompetence or emotional instability) and moral (e.g., dishonesty or laziness) in contrast to social scientists who emphasize amoral and structural factors. Part of what is at issue here is the social scientist's sensitivity to the impact of organizational structure on the person who occupies a particular position. Insiders, on the other hand, tend to explain organizational behavior in terms of the personal characteristics of the individuals who hold positions. Also involved is the evaluative researcher's more secularized explanation of human behavior which leads him to emphasize factors outside the realm of free choice.

Academic versus practical experience. Because the evaluative researcher typically approaches social action from the perspective of an academic discipline, his knowledge of practical affairs is usually highly incomplete. Unless he has had administrative experience in an action setting, the evaluative researcher is not likely to comprehend fully the administrator's position. Political constraints, budgetary problems, and limitations of personnel and facilities are among the realities which an evaluative researcher, preoccupied with the substance of programs, may underestimate. Similarly, it is difficult for administrators with limited research training to understand the evaluative researcher's emphasis on methodology.

Client Activism

Much of the innovative social programming in recent years has been directed at a reduction of the incidence of poverty. At the same time, group self-consciousness has been growing among the minorities who represent a substantial portion of the poor. Stimulated by the civil rights movement and by professional community developers, minority activists have taken a significant interest in local community affairs—including the social programs directed at the poor. As clients or spokesmen for clients of anti-poverty programs, activists have pressed for extensive participation if not full control of these programs by clients, at the levels of both policy and implementation. Anti-poverty programs consequently have often been sur-

rounded by substantial and continuous conflict over such matters as representation on and authority of boards, employment policies and practices, and the substance and administration of programs. Beyond the direct programming implications of minority activism, the movement has added to the challenges with which the evaluative researcher is confronted.

Even though evaluative researchers may firmly believe that their efforts contribute ultimately to the cause of the poor, minority activists may confront them with great hostility. Part of what is involved is that the basic issues which strain evaluator-administrator relations even more thoroughly set evaluative researchers apart from low-income program clients. Preoccupied with the immediate, tangible, dramatic, and personal, the minority activist is likely to be impatient with the evaluator's concern with the future, abstract concepts, orderly procedures, and impersonal forces. In contrast to the activist who often seeks to generate open conflict, the evaluative researcher typically emphasizes cooperative approaches to problem solving. The evaluator may also find himself in an awkward position in the power struggle between client spokesmen and professional administrators. If he entered the program at the invitation of a funding agency or a professional administrator, the evaluative researcher is likely to be mistrusted immediately by minority activists who see him as a potential spy. Indeed, if evaluative criteria are limited to those acceptable to administrators, and if evaluation findings are subject to administrative review prior to being publicized, client spokesmen have good reason to challenge the evaluator's contribution.

Some of the minority activist's hostility toward the evaluator is attributable to a general antipathy toward social research. A wide-spread complaint of minority spokesmen is that they have been "surveyed to death." Perhaps social research has come to symbolize for some the powerlessness of the poor. Resentment is obvious in the poverty spokesman's view that social research on poverty has nearly always been initiated by outsiders and addressed to issues defined by outsiders. The cooperation of the poor has been solicited with rhetoric which links research to desired social goals; yet, it is difficult for the poor to see tangible benefits stemming from social research. In fact, many activists cynically view research as a substitute for needed action. General antagonism toward social research is also linked to the activist's political ambitions. The independent social scientist who does poverty research is a potential competitor for the activist who would like to control the flow of information from poverty areas. The would-be indigenous spokesman from the poor has reason to be anxious if his claims are challenged by respected social scientists.

Client representatives are additionally justified in challenging evaluative researchers if they have reason to question the latter's assurances of confidentiality in the use of information about persons. Walsh (1969)

reported an incident that developed in the evaluation of an Office of Economic Opportunity project concerned with delinquent gangs. After confidentiality had been pledged and significant information on individuals had been collected, the study group complied (however reluctantly) with a Senate committee's subpoena of raw data.

Strategies for Establishing and Maintaining the Evaluation Role

A number of experienced evaluators have suggested strategies for dealing with the problems which can be expected in establishing and maintaining the evaluation role.

A basic administrative issue is the comparative advantages and disadvantages of "inside" and "outside" evaluators. The "inside" evaluator is a staff member of the organization whose programs are evaluated; the "outside" evaluator is a consultant from outside the organization. The following are some of the arguments which have been presented in favor of outsiders: (a) they tend to be better able to maintain their objectivity; (b) they are more likely to be able to include evaluative criteria which question basic organizational premises; (c) they may be able to mediate more effectively if there is extensive internal conflict; (d) they usually are better protected from problems of marginality and status incongruity; (e) they are better able to avoid unwelcome nonresearch tasks. It has been suggested that insiders have the following advantages: (a) they are usually able to develop a more detailed knowledge of the organization and its programs; (b) they are in a better position to do continuing research. Likert and Lippitt (1953), Weinberger (1969), Weiss (1966), McEwan (1956), and Rodman and Kolodny (1964) are among those who have addressed themselves to these arguments. Luchterland (1967), however, pointed out that outsiders cannot always be counted on to be more objective than insiders. When they are concerned with maintaining good relations with clients, outsiders may slant their interpretations to accommodate their clients' interests; alienated inside evaluators may be inclined to report on their agencies' programs with stark objectivity. Yet, funding agencies, spokesmen for clients, and the general public usually consider the reports of external evaluators more credible. As Lortie (1967) pointed out, persons and organizations cannot be trusted to act as judges in their own cases. Their self-appraisals cannot be accepted without question. When evaluation is conducted for the purpose of accounting to an outside body, utilization of external evaluators appears preferable. If, on the other hand, evaluation is conducted to assist an organization in its program development efforts, an internal evaluation unit may be able to contribute more effectively.

If the evaluative researcher hopes to contribute to internal program development, it has been suggested that he take early steps to establish effective ties with those who make key programming decisions. Sensitivity

to the locus of decision making is important. Relations with administrators are always important; but in more decentralized and democratic organizations, evaluative researchers may find it appropriate to work more closely with the professional practitioners (e.g., physicians, social workers, teachers) who are more concerned with the substance of programs.

Some authors emphasized the importance of the evaluator's organizational position (see Argyris, 1958; Bennis, 1965; Rosenthal & Weiss, 1966; Suchman, 1967a; Whyte & Hamilton, 1964). The evaluator's prestige and power are considered to be positively related to the likelihood that his findings will be implemented. If the evaluator is an insider, it is important that he have a prestigious position within the organization. Similarly, if he is an outsider, it is helpful if he has strong professional and organizational credentials. It is also important for an outside evaluator to be associated with someone of high status in the action organization—a relationship which Sussman (1966) called the "Merlin role." However, when he makes status claims, the evaluator must concern himself with the possible resentment of staff subordinates. If they believe he receives more status prerogatives than he deserves, staff subordinates may withhold their cooperation from the evaluation effort.

As he begins working with agency representatives, it is important for the evaluator to create what Likert and Lippitt (1953) called an "image of potential." The evaluator must, for example, provide administrators and practitioners with assurance of his technical competence, his understanding of the action setting, and his personal integrity and decency (Warren, 1963).

A mutual clarification of expectations at an early stage in the relationship would be useful. Administrators, for example, should be informed of some of the limitations of the contribution of evaluative research. Evaluators might need to explain that their work cannot resolve fundamental value issues nor can it, by itself, resolve deep-seated conflicts between administrators and their staff or between the agency and its clients. If evaluation is to be used for program development purposes, evaluators should attempt to gauge the extent to which policy makers tolerate challenges to their basic premises. An early agreement regarding the manner in which evaluation results will be publicized should also be sought. If the purpose of the evaluation is summative and it is externally sponsored, advance agreement is desirable on the extent to which persons and organization units will be identified in published reports. The evaluative researcher's interest in pursuing professional research interests should also be discussed. For his work to be relevant in the action setting, the evaluative researcher may have to postpone the pursuit of some of his personal intellectual interests. It may be desirable for him to reach an early and explicit agreement with the funding agency and program administrators on the extent to which he is free to use his time and project data for professional research purposes.

The evaluative researcher should inform himself not only about available action alternatives, but about the timing of decision making. If results are to be used, evaluation must be addressed to pertinent issues and results must be available when needed.

Evaluators may be able to make a greater contribution if they can modify the policy maker's approach to programming. Sadofsky (1966) suggested that the program operator's fear of failure might be diminished if action projects were considered as experiments. Failure, then could be seen as a learning opportunity. Weiss (1966) recommended that instead of judging programs in simple success or failure terms, the administrator should be encouraged to ask questions about the relative effectiveness of alternative programs.

It is generally suggested that evaluators work closely with administrators in establishing evaluative criteria so that evaluators may become more fully aware of administrative concerns and so that administrators may become more committed to the evaluation process (Freeman & Sherwood, 1965). Collaboration in the identification of criteria or goals may help evaluators base their work on variables more explicit, realistic, and perhaps more comprehensive than the objectives shown in official program documents. Stake (1967), however, introduced a note of caution. He argued that administrators or practitioners should not be expected to work at the high level of abstraction required for the writing of behavioral goals. Rather, evaluators should draft statements of objectives which attempt to reflect and clarify the intent of administrators. Coleman (1969) similarly pointed out that because administrators are often not fully aware of their decision-making criteria, evaluative researchers themselves may have to discover these criteria.

A number of writers pointed to a need for evaluative researchers to consider a wide variety of potential program effects, including those which are unintended and undesired. Scriven (1969) emphasized the evaluator's responsibility to focus his efforts on appropriate evaluative criteria. Campbell (1969), concerned with undesired side effects, recommended that several outcome measures be utilized including those proposed by "loyal opponents." However, because of limited evaluation budgets and the relatively narrow range of alternatives which the administrator sees as open, the evaluator often finds it prudent to narrow the range of his inquiry. Aware of his lack of power, but hopeful of being able to influence policy makers within a limited but significant range of decision alternatives, the evaluator might find it desirable to ignore some potential evaluative criteria.

It is important for the evaluator to take some steps to obtain cooperation not only from administrators, but also from staff members who carry out programs. Staff support is critical if programs are to be carried out as designed and if program records, essential for evaluation purposes, are to

be maintained. Staff cooperation, however, cannot be taken for granted. A basic problem here is that the evaluator's relationship with top administrators puts him in the same organizational position as an inspector or policeman. If he hopes to obtain staff cooperation, the evaluator must insist that program evaluation is quite different from the evaluation of individuals or organizational units. Thus, Likert and Lippitt (1953, p. 161) emphasized that staff members must be assured

that the objective of research is to discover the relative effectiveness of different methods and principles and that the study is in no way an attempt to perform a policing function. The emphasis must be on discovering what principles work best and why, and not on finding and reporting which individuals are doing their jobs well or poorly.

Staff subordinates must, then, be given emphatic assurance of confidentiality and anonymity. It is also desirable to obtain a commitment from administrators to share evaluation findings openly with subordinates. If evaluation will add to the record-keeping duties of practitioners, evaluators may be wise to provide practitioners with added compensation or staff support.

Because of pressure to produce results quickly, timing may be a critical concern in the organization of evaluation efforts. Time pressures, must, of course, be given strong consideration in the selection of a methodological strategy. Grobman (1968) suggested that evaluators use a formal planning procedure such as Program Evaluation and Review Technique (PERT) to assure that evaluation work is completed within a tight time schedule. In some cases evaluators may wish to report interim findings, either to aid in an immediate decision problem or to keep administrators interested in the evaluation process. Early feedback, however, is a problem for evaluators if it causes administrators to substantially change programs before enough cases have been observed to satisfy the requirements of an experimental design.

Utilization of evaluation findings may depend on the manner in which results are reported. Some authors pointed to the need for clear, concise, and even dramatic presentation of findings. Sadofsky (1966) warned that delivering results to an administrator publicly, without warning, may produce a defensive reaction to findings. It may be desirable to supplement written reports with personal meetings with administrators. Mann and Likert (1952) recommended a series of small group meetings both at top administrative levels and through the ranks of subordinates to facilitate communication of results and to stimulate interests in following through on the action implications of results. They argued that the pressures generated in small groups increase commitment to implementation of recom-

mended changes. Argyris (1958) proposed asking administrators and practitioners for their own diagnoses first to reduce the likelihood that they reject research findings as too obvious.

Strategies for Obtaining Cooperation from Client Spokesmen

Where clients' cooperation with evaluators may be a problem, support of client spokesmen should be sought at an early stage. Funding agencies or administrators should initially explain the rationale for evaluation and the allocation of evaluation funds. Participation of client spokesmen in the selection of an evaluator may also be advisable. Since employment opportunity is a central concern among minority activists, it is desirable for evaluators to employ some members of the population served by programs. (Such a commitment may make it necessary for the evaluator to place more emphasis on staff training and supervision than he would otherwise.) Even more than staff subordinates, client spokesmen need assurance that confidential personal information will be used only for overall evaluation purposes. They also need to be convinced that, unlike basic research, evaluative research is designed to have immediate action implications. Client spokesmen need assurance that evaluation results will be available to them and that they will have full opportunity to participate in their interpretation. An evaluator may be able to satisfy some of the personalistic concerns of poverty groups by spending enough time with minority spokesmen so that they know and trust him as an individual.

In some situations the level of conflict between client spokesmen and established agencies may be so great that cooperation in program evaluation is not a realistic possibility. In these cases it may be preferable for each group to sponsor its own evaluation enterprise. Funding agencies may find it advisable in these cases to provide organized client spokesmen with the funds needed for their independent evaluation of programs.

Methodological Considerations

The methodological principles which apply to the evaluation of social programs are not different from those of general behavior science inquiry. There is some regularity, however, in the problems of measurement and design which arise in evaluative research.

Texts and manuals on the methodology of evaluative research include those written by Hayes (1959), Herzog (1959), Fairweather (1967), Suchman (1967a), and Grobman (1968). Writings on field experiments by such persons as French (1953), Campbell (1957, 1967, 1969), Campbell and Stanley (1963), and Barnes (1967) are also highly relevant for evaluative research. Among those who have written extensively about methodological problems in evaluation work are Hyman, Wright, and Hopkins

(1962), Cronbach (1963), Whyte and Hamilton (1964), Mann (1965), Greenberg (1968), Lerman (1968), Scriven (1967), Stake (1967) and Glass (1971).

Measurement

A basic step in evaluation is the identification of objectives and their measurement. Suchman (1966) suggested that formulated objectives have five aspects: (a) the content of the objective (i.e., the social conditions or behavior patterns to be changed by the program), (b) the target of the program (i.e., the population to which the program is addressed), (c) the time within which the change is to take place, (d) the number of objectives (if they are multiple), and (e) the extent of expected effect. Freeman (1965), Suchman (1967a), Greenberg (1968), and Weiss (1966) urged a distinction among immediate, intermediate, and ultimate objectives. Measurements focused on immediate and intermediate objectives are particularly important when evaluation results are needed before ultimate objectives can be realized. If immediate and intermediate objectives are used as substitutes for ultimate objectives, however, the burden is on the evaluator to argue the validity of the hypothesized links to ultimate objectives. When programs fail to realize ultimate objectives, utilization of a hierarchy of objectives may also be useful in accounting for their limited success.

Because the realities of program operations are often inconsistent with public project descriptions, measurement of program inputs has also been recommended. Greenberg (1968) termed observation of administrative patterns and analysis of service statistics as "quasi-evaluation." Coleman (1969) urged a distinction between resources allocated by organizations and services actually received by clients. Analysis of these administrative data may be useful for preliminary program screening purposes. To the extent that organizations are unable to deliver services to clients, expectations of program effectiveness are, of course, diminished.

As discussed previously, it is also desirable for the evaluator to anticipate and measure possible unintended effects of programs including those which are undesirable. Scriven (1967) recommended that evaluators consider secondary effects of programs such as impact on the individuals and organizations who conduct programs and impact on those who regularly interact with program beneficiaries.

Identification of variables is only a first step in the measurement process. Evaluators are often confronted with serious obstacles in seeking the valid, reliable, and sensitive measures they need. Lerman (1968) and Campbell (1969) pointed to the shortcomings of the agency records on which evaluators are often dependent. When he uses agency records, the evaluator must take into account the fact that these data may reflect the organi-

zational, professional, and individual interests of those who maintain the records as much as they reflect the behavior which they are supposed to measure.

Because of his refined information requirements and the poor quality of agency records, the evaluator frequently must collect original data. When he gathers his own data, the evaluator is faced with additional problems. Data collection may add enormously to the cost of evaluation. Administrators and practitioners may object that it interferes with their programming efforts either because it takes away from the time available for programming or because it may jeopardize client or community acceptance of the program. Evaluators may be concerned that through their data collection activities they may enhance client awareness of the program, thereby adding artificially to its apparent or actual effectiveness. The evaluator may cope with these data collection problems by using unobtrusive measures (Webb et al., 1966) or by disguising the relationship between his data collection and the program (Seashore, 1964). He may also address these problems through his selection of a research design. Campbell (1957) suggested use of the Solomon four-group design or a design requiring only posttest measurements. Also see Suchman (1967a) and Wuebben (1968). Although Hyman, Wright, and Hopkins (1962) reviewed evidence indicating that the sensitizing or practice effects of pretesting are often negligible, the evaluator is clearly advised to take steps to guard against this potential source of measurement error.

Freeman (1963) urged that evaluators use behavioral rather than attitudinal measures of program objectives because policy makers are more likely to be impressed with behavioral data. Deutscher (1969) similarly argued in favor of direct behavioral measures because they pose fewer validity problems than do procedures designed to provide estimates of hypothetical behavior.

Beyond the sensitizing effects of measurement, widespread awareness of evaluative criteria and measurement procedures can have important undesired effects on the ways in which programs are administered and interpreted by clients. The danger is that administrative units, practitioners, or clients may artificially redirect their behavior to affect the outcome of evaluation. The problem is particularly acute when incomplete sets of evaluative criteria and imperfect measures are used to judge the performance of participants. Considerable attention has been given to this problem in higher education, where it has been argued that faculty are often excessively concerned with numbers of publications and students are overly preoccupied with grades. By emphasizing their concern with program concepts rather than specific participants, evaluative researchers may be able to deal with this problem with some effectiveness.

Timing of measurement is another serious issue in evaluative research (Freeman & Sherwood, 1965; Hyman & Wright, 1967; Harris, 1963). It is often not clear how soon program effects can be expected. The stability and durability of changes brought about by programs may also be in doubt. Ideally, the problem is addressed through continuous or at least repeated measurement of output variables. Many evaluative researchers, however, find themselves in situations where they have an opportunity only for a single post-treatment measurement. The timing of such a measurement may have most important implications for the outgrowth of evaluative research.

Design

To assure that changes in measured behavior can be attributed exclusively to the program at hand, evaluative researchers prefer to employ some form of an experimental design. From an evaluation perspective, it is desirable that clients be assigned randomly to treatment and control groups. Adequate control, however, is difficult to achieve in an action setting. Suchman (1967b) cited two obstacles to the effective use of control groups: (a) service orientation—administrators, practitioners, and client representatives are reluctant to allow services to be withheld from those who might benefit from them; (b) self-selection—it is difficult to either refuse service to those who seek it or provide service to those who resist it. Mann (1965) further observed that in an organizational setting, innovative approaches may “spread like a disease” to control groups. In discussing the evaluation of community-wide programs, Greenberg (1968) pointed out the added problem of finding truly equivalent communities. Where control groups are not possible, experimental control may be approximated through some design adjustments. One approach is to match participants with nonparticipants and compare them through the use of analysis of covariance. The time-series design (Hyman, Wright, & Hopkins, 1962; Campbell & Stanley, 1963; Campbell, 1969; Gottman, McFall, & Barnett, 1969) is an alternative by which the treatment group is used as its own control through repeated measurements of outcome variables beginning well before program implementation. For treatment of further design possibilities, see Campbell and Stanley (1963) and Campbell (1969).

Lerman (1968) argued that evaluators should resist the common administrative assumption that evaluation is based on those who complete treatments. Rather, evaluation should be based on the population in need of services. Lerman pointed out that the issue is particularly critical among private agencies which can be selective in whom they accept as clients.

It may be possible to use comparison groups in action settings where control groups are unacceptable. Unlike the control group which receives no treatment, the comparison group receives an alternate treatment. Where

policy makers are committed to the principle of providing additional services, a comparison-group design may actually provide more useful information than a design using only a strict control.

Social programs usually are not expected to produce a dramatic impact. If evaluation is to document subtle but important changes, large samples or highly sensitive designs are necessary (Freeman, 1963). The conservatism regarding rejection of null hypotheses which often prevails in academic research may also be inappropriate in the formulation of decision criteria for evaluative research. Rather, evaluators may wish to be cautious in drawing negative conclusions regarding innovative programs (Miller, 1965b).

A persistent problem in the design of evaluative research is the separation of effects of program content from effects of practitioners' characteristics. Staff enthusiasm and confidence may be critical variables in innovative programs. Design adjustments are particularly difficult when the number of practitioners is small. Greenberg (1968) suggested that program personnel be rotated between treatment and control conditions. Some of Rosenthal's (1966) suggestions for controlling experimenter expectancy effects in social psychological research appear to be applicable. Special training and supervision of practitioners may be introduced to reduce variability in practitioner behavior. Alternately, it may be possible to conduct some programs with minimal practitioner-client contact.

Program recipients sometimes contribute to the effectiveness of a program through their feelings of self-importance as persons selected for special attention (Hawthorne effect) or through their faith in the program (placebo effect). The impact of the Hawthorne or placebo effects is likely to be particularly great when the program is new and experimental and the participants are volunteers. Scriven (1967) suggested the use of multiple experimental groups to separate these effects from those of programs. He urged that enthusiasm be held constant while treatments are varied. Trow (1967), however, pointed out that some administrators may try to capitalize on Hawthorne effects by attempting to build an experimental climate into their normal programming. Sommer (1968) similarly argued that the Hawthorne effect is not an extraneous disruptive influence; rather it is an important and ever present factor in any field situation. "Environmental changes do not act directly upon human organisms. They are interpreted according to the individual's needs, set, and state of awareness [p. 594]." The implication seems to be that if the effects of social programs are to be fully understood, it is important that the client population's predisposition toward and interpretation of programs be an integral part of comprehensive evaluative research.

New programs often pose difficulties for evaluators which are not present in the case of well established programs. On the one hand, the evaluative researcher must be prepared to deal with the positive effects of novelty,

special attention, and enthusiasm. On the other hand, he must look for some of the strictly administrative problems common in the implementation of a new program which can account for the failure of an otherwise soundly conceived program (Hyman & Wright, 1967). In the case of innovative programs, it is particularly important that administrators be free to modify their procedures on the basis of their early experiences in implementing the program (Marris & Rein, 1967). These modifications pose an enormous problem for evaluation if research designs call for a lengthy commitment to a highly specific set of procedures. If, as Glass (1971) recommended, evaluators focus on program concepts rather than specific procedures, their experimental designs may be able to accommodate procedural adjustments as long as basic concepts of the program remain intact.

Program outcomes may also be affected by many other variables which cannot be controlled in a single evaluative study. Among these variables are the physical characteristics of the program site and the duration and intensity of the program.

Because action programs are often ineffective and because experimental evaluation is often very expensive, Rossi (1967) recommended a two-phased approach to evaluation. First, correlational designs would be used to identify promising programs. Then, controlled experiments would be conducted to evaluate the relative effectiveness of those programs which passed the initial screening.

Rigid evaluation designs are most easily implemented in the programs conducted by highly centralized organizations having extensive control over their clients. Prisons, hospitals, and residential schools are among the organizations most likely to have these characteristics. When programs involve a number of autonomous organizations, are conducted by practitioners with considerable personal and professional autonomy, and are directed at client populations whose willingness to cooperate is highly uncertain, evaluators often must be satisfied to use limited methodological tools. Effective programming is, of course, also very difficult under these circumstances.

Contemporary community-wide anti-poverty programs are among those in which it is most difficult for evaluators to use highly controlled experimental designs. The relative contribution of various components of these large-scale programs may be difficult to determine because of clients' uncontrolled exposure to several programs. It may also be difficult to determine the extent to which new programs are supplements to rather than substitutes for earlier programs. Weiss and Rein (1969) further noted that in the case of highly diffuse and unstable programs, it is particularly difficult to select and operationalize evaluative criteria which are sufficiently broad in scope to reflect a program's full range of consequences—especially consequences which are unintended.

In these settings the evaluator must look for research strategies which are realistic and, at the same time, yield a maximum of useful information. Particularly in the case of completely innovative programs for which evaluation results are needed at an early stage, informal approaches usually associated with exploratory research may be most appropriate. Observational techniques and informal interviewing can often provide more useful and rapid feedback than can formal experimentation (Weiss & Rein, 1969). Lazarsfeld, Sewell, and Wilensky (1967) observed that because the decision process in these programs is continuous, evaluation must take place at many points. They recommended concurrent evaluation, a procedure by which records are kept of all decisions, including information on rejected alternatives and expected outcomes. Perhaps, as Benedict et al. (1967) suggested, what is needed is evaluation which combines rigorous experimental data with a "natural history" account of events and actors before, during, and after program implementation.

Decision makers are usually concerned with efficiency as much as they are with effects of programs. Evaluators, therefore, should be prepared to deal with the relationship between cost and effectiveness. In some cases cost analysis is straightforward; in others, it adds another complex dimension to evaluation.

Concluding Thoughts

Clearly, evaluative research is an activity surrounded by serious obstacles. Satisfied with informal and impressionistic approaches to evaluation, policy makers are often reluctant to make the investment needed to obtain verifiable data on the effects of their programs. Evaluative researchers are typically confronted with problems of measurement and design which greatly restrict their ability to reach unambiguous conclusions. Abrasive relations with practitioners and clients can add to the evaluator's difficulties in obtaining information. Evaluative research is often addressed to a distressingly narrow range of issues; results are not as fully or widely disclosed as they might be; highly pertinent findings are often ignored by policy makers. It is little wonder that many social scientists regard evaluative research as a dubious enterprise.

Yet, the argument for emphasizing evaluative research in social programming is strong. Expenditures in this country for social service programs (including health and education) are enormous. Yet there is reason to be dissatisfied with the effectiveness of many of these programs. Increases in program costs tend to be much more conspicuous than improvements in the quality of services. If it is agreed that social programs should be strengthened and that improvement is most likely to come about through the use of rational methods, it is clear that the evaluation role is important and should be emphasized. The often subtle results of social programs require

the methods of empirical research to obtain precise information on program effectiveness.

Evaluative researchers can take a number of steps to improve their contributions to program development. They can become more skillful in applying their methodological tools to specific evaluation problems. They can become more knowledgeable about the decision problems of action organizations, and thus recommend more appropriate evaluation strategies. Greater personal familiarity with action settings may make evaluators more effective in working with practitioners and clients. The climate for evaluation might be improved if evaluators were to place more emphasis on educating administrators, practitioners, and client representatives regarding the role of evaluation in program development. Evaluators might develop more effective ways of communicating the action implications of their findings. Behavioral scientists who assume administrative roles in programs can also help by showing how programs can be structured to accommodate evaluation requirements.

If, however, evaluative research is to make its full contribution, substantial changes must be made in society's overall approach to social programming. Legislators and other public officials, reflecting widespread public concern, must significantly raise their demands for the effectiveness and efficiency of programs. In addition, they must learn to focus more on program goals so that they can assume a more experimental attitude toward specific programming strategies (Campbell, 1969). Such fundamental changes in orientation toward social programming would lead to greatly expanded interest in evaluative research. If there were a more serious emphasis on performance standards and on the search for more effective program approaches, evaluative researchers would be more often able to obtain the political and administrative support needed to employ experimental designs. Behavioral scientists who hope to contribute to the effectiveness of social programs through evaluative research need to concern themselves not only with immediate methodological and organizational problems but with the larger issues concerning the social context in which social programs are conducted.

References

- Argyris, C. Creating effective relationships in organizations. *Human Organization*, 1958, 17 (1), 34-40. (Republished: In R. Adams & J. Preiss (Eds.), *Human Organization Research*. Homewood, Ill.: Dorsey, 1960.)
- Barnes, L. Organizational change and field experiment methods. In V. Vroom (Ed.), *Methods of organization research*. Pittsburgh, Penna.: University of Pittsburgh Press, 1967.
- Benedict, B. A., Calder, P. H., Callahan, D. M., Hornstein, H. A., & Miles, M. B. The clinical-experimental approach to assessing organizational change efforts. *Journal of*

- Applied Behavioral Science*, 1967, 3, 347-380.
- Bennis, W. Theory and method in applying behavioral science to planned organizational change. *Journal of Applied Behavioral Science*, 1965, 1, 337-360.
- Brooks, M. The community action program as a setting for applied research. *Journal of Social Issues*, 1965, 21, 29-40.
- Bynder, H. Sociology in a hospital: A case study in frustration. In A. Shostak (Ed.), *Sociology in action*. Homewood, Ill.: Dorsey, 1966.
- Campbell, D. T. Validity of experiments in social settings. *Psychological Bulletin*, 1957, 54, 297-312.
- Campbell, D. T. Administrative experimentation, institutional records, and nonreactive measures. In J. Stanley (Ed.), *Improving experimental design and statistical analysis*. Chicago: Rand McNally, 1967.
- Campbell, D. T. Reforms as experiments. *American Psychologist*, 1969, 24, 409-429.
- Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963. (Republished: *Experimental and quasi-experimental design for research*. Chicago: Rand McNally, 1966.)
- Cherns, A. Social research and its diffusion. *Human Relations*, 1969, 29, 209-218.
- Coleman, J. S. Evaluating educational programs: A symposium. *The Urban Review*, 1969, 3 (4), 6-8.
- Cronbach, L. J. Course improvement through evaluation. *Teachers' College Record*, 1963, 64, 672-683.
- Deutscher, I. Looking backward: Case studies in the progress of methodology in sociological research. *American Sociologist*, 1969, 4, 35-41.
- Downs, A. Some thoughts on giving people economic advice. *American Behavioral Scientist*, 1965, 9 (1), 30-32.
- Etzioni, A. Two approaches to organizational analysis: A critique and a suggestion. *Administrative Science Quarterly*, 1960, 5, 257-278.
- Fairweather, G. *Methods of experimental social innovation*. New York: Wiley, 1967.
- Ferman, L. A. Some perspectives on evaluating social welfare programs. *Annals of the American Academy of Political and Social Science*, 1969, 385, 143-156.
- Freeman, H. E. Strategy of social policy research. In H. E. Freeman (Ed.), *Social welfare forum*. New York: Columbia University Press, 1963.
- Freeman, H. E., & Sherwood, C. Research in large scale intervention programs. *Journal of Social Issues*, 1965, 21, 11-28.
- French J. Experiments in field settings. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Holt, 1953.
- Glass, G. V. The growth of evaluation methodology. *AERA Monograph Series on Curriculum Evaluation*, No. 7, 1971, in press.
- Gottman, J. M., McFall, R. M., & Barnett, J. T. Design and analysis of research using time series. *Psychological Bulletin*, 1969, 72, 299-306.
- Greenberg, B. G. Evaluation of social programs. *Review of the International Statistical Institute*, 1968, 36, 260-277.
- Grobman, H. *Evaluation activities of curriculum projects: A starting point*. Chicago: Rand McNally, 1968.
- Hall, R. The applied sociologist and organizational sociology. In A. Shostak (Ed.), *Sociology in action*. Homewood, Ill.: Dorsey, 1966.
- Harris, C. W. *Problems in measuring change*. Madison: University of Wisconsin, 1963.
- Hayes, S. P. *Measuring the results of development projects*. New York: UNESCO, 1959.
- Herzog, E. *Some guidelines for evaluation research*. Washington: U.S. Government Printing Office, 1959.
- Horowitz, I. The academy and the polity: Interaction between social scientists and federal administrators. *Journal of Applied Behavioral Science*, 1969, 5, 309-335.
- Hyman, H., & Wright, C. Evaluating social action programs. In P. Lazarsfeld, W. Sewell, & H. Wilensky (Eds.), *The uses of sociology*. New York: Basic Books, 1967.
- Hyman, H., Wright, C., & Hopkins, T. *Applications of methods of evaluation*. Berkeley: University of California, 1962.
- Lazarsfeld, P., Sewell, W., & Wilensky, H. (Eds.) Introduction to *The uses of sociology*. New York: Basic Books, 1967.

- Lerman P. Evaluative studies of institutions for delinquents: Implications for research and social policy. *Social Work*, 1968, 13 (3), 55-64.
- Likert, R., & Lippitt, R. Utilization of social science. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Holt, 1953.
- Lippitt, R., Watson, J., & Westley, B. *The dynamics of planned change*. New York: Harcourt Brace, 1958.
- Lortie, D. C. The cracked cake of educational custom and emerging issues in evaluation. Paper presented to the Symposium on Problems in the Evaluation of Instruction, University of California, Los Angeles, December 1967.
- Luchterhand, E. Research and the dilemmas in developing social programs. In P. Lazarsfeld, W. Sewell, & H. Wilensky (Eds.), *The uses of sociology*. New York: Basic Books, 1967.
- Mann, F., & Likert, R. The need for research on the communication of research results. *Human Organization*, 1952, 11 (4), 12-13.
- Mann, J. *Changing human behavior*. New York: Scribner's, 1965.
- Mann, J. Evaluating educational programs: A symposium. *The Urban Review*, 1969, 3 (4), 12-13.
- Marris, P., & Rein, M. *Dilemmas of social reform*. New York: Atherton Press, 1967.
- McEwen, W. J. Position conflict and professional orientation in a research organization. *Administrative Science Quarterly*, 1956, 1, 208-224.
- McKelvey, W. Expectational non-complementarity and style of interaction between professional and organization. *Administrative Science Quarterly*, 1969, 14 (1), 21-32.
- Merton, R. Role of the intellectual in public bureaucracy. In R. Merton, *Social theory and social structure*. New York: Free Press, 1957.
- Miller, S. M. Evaluating social action programs. *Trans-action*, 1965, 2 (3), 38-39. (a)
- Miller, S. M. Prospects: The applied sociology of the center-city. In A. Gouldner & S. M. Miller (Eds.), *Applied sociology*. New York: Free Press, 1965. (b)
- Perry, S. E., & Wynne, L. Role conflict, role redefinition, and social changes in a clinical research organization. *Social Forces*, 1959, 38, 62-65.
- Rein, M., & Miller, S. M. The demonstration project as a strategy of change. In M. N. Zald (Ed.), *Organizing for community welfare*. Chicago: Quadrangle Books, 1967.
- Rodman, H., & Kolodny, R. Organizational strains in the researcher-practitioner relationship. *Human Organization*, 1964, 23, 171-182. (Republished: In A. Gouldner & S. M. Miller (Eds.), *Applied sociology*. New York: Free Press, 1965.)
- Rosenthal, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.
- Rosenthal, R., & Weiss, R. Problems of organizational feedback. In R. Bauer (Ed.), *Social indicators*. Cambridge, Mass.: MIT Press, 1966.
- Rossi, P. Evaluating social action programs. *Trans-action*, 1967, 4, 51-53.
- Rossi, P. Evaluating educational programs: A symposium. *The Urban Review*, 1969, 3 (4), 17-18.
- Sadofsky, S. Utilization of evaluation results: Feedback into the action program. In J. Shmelzer (Ed.), *Learning in action*. Washington: U.S. Government Printing Office, 1966.
- Schulberg, H., & Baker, F. Program evaluation models and the implementation of research findings. *American Journal of Public Health*, 1968, 58, 1248-1255.
- Scriven, M. The methodology of evaluation. *AERA Monograph Series on Curriculum Evaluation*, No. 1, *Perspectives of curriculum evaluation*, 1967.
- Scriven, M. Evaluating educational programs: A symposium. *The Urban Review*, 1969, 3 (4), 20-22.
- Seashore, S. Field experiments with formal organizations. *Human Organization*, 1964, 23, 164-170.
- Shepard, H. A. Nine dilemmas in industrial research. *Administrative Science Quarterly*, 1956, 1, 295-309.
- Smith, J., Sim, F., & Bealer, R. Client structure and the research process. In R. Adams & J. Preiss (Eds.), *Human organization research*. Homewood, Ill.: Dorsey, 1960.
- Sommer, R. Hawthorne dogma. *Psychological Bulletin*, 1968, 70, 592-595.
- Stake, R. The countenance of educational evaluation. *Teachers' College Record*, 1967, 68, 523-540.

- Suchman, E. A model for research and evaluation on rehabilitation. In M. Sussman (Ed.), *Sociology and rehabilitation*. Washington: American Sociological Association, 1966.
- Suchman, E. *Evaluative research*. New York: Russel Sage Foundation, 1967. (a)
- Suchman, E. Principles and practices of evaluative research. In J. Doby (Ed.), *An introduction to sociological research*. (2nd ed.) New York: Appleton-Century-Crofts, 1967. (b)
- Suchman, E. Evaluating educational programs: A symposium. *The Urban Review*, 1969, 3 (4), 15-17.
- Sussman, M. The sociologist as a tool of social action. In A. Shostak (Ed.), *Sociology in action*. Homewood, Ill.: Dorsey, 1966.
- Trow, M. Methodological problems in the evaluation of innovation. Paper presented to the Symposium on Problems in the Evaluation of Instruction, University of California, Los Angeles, December 1967.
- Walsh, J. Anti-poverty R & D: Chicago debacle suggests pitfalls facing OEO. *Science*, 1969, 165, 1243-1245.
- Warren, R. *Social research consultation*. New York: Russell Sage Foundation, 1963.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.
- Weinberger, M. Evaluative action programs: Observations by a market researcher. *The Urban Review*, 1969, 3 (4), 23-26.
- Weiss, C. H. Planning an action project evaluation. In J. Schmelzer (Ed.), *Learning in action*. Washington: U. S. Government Printing Office, 1966.
- Weiss, C. H. Utilization of evaluation: Toward comparative study. Paper presented to the American Sociological Association, September 1966, Miami Beach. (Printed in *The use of social research in federal domestic programs*, Part III, *The relation of private social scientists to federal programs on national social problems*. Washington: U.S. Government Printing Office, 1967.)
- Weiss, R., & Rein, M. The evaluation of broad-aim programs: A cautionary case and a moral. *Annals of the American Academy of Political and Social Science*, 1969, 385, 133-142.
- Whyte, W. F., & Hamilton, E. *Action research for management*. Homewood, Ill.: Dorsey, 1964.
- Wrightstone, J. W. Evaluating educational programs: A symposium. *The Urban Review*, 1969, 3 (4), 5-6.
- Wuebben, P. Experimental design, measurement, and human subjects: A neglected problem of control. *Sociometry*, 1968, 31, 89-101.
- Wynne, E. Evaluating education programs: A symposium. *The Urban Review*, 1969, 3 (4), 19-20.

AUTHOR

FRANCIS G. CARO *Address*: Florence Heller Graduate School for Advanced Studies in Social Welfare, Brandeis University *Title*: Associate Professor *Degrees*: B.S., Marquette University; Ph.D., University of Minnesota *Specialization*: Evaluative research.