

*TRACES are what evaluators left behind—discoveries, records, tracks—which made marks on the profession of program evaluation. Published here are excerpts from our past (e.g., articles, passages from books, speeches) that show where we have been or affected where we are g(r)o(w)ing. Suggestions for inclusion should be sent to the Editor, along with rationale for their import. Photocopies of the original printed versions are preferred with full bibliographic information. Copying rights will be secured by the Editor.*

**Editor's Note:** I am delighted to be able to reprint Carol H. Weiss' 1973 paper entitled "Where politics and evaluation research meet." Many years ago when I first became interested in evaluation, her work made sense to me because of her obvious appreciation for the realities of assessing (and implementing) people-type programs. But that, in itself, would not suffice for honoring her here in *Traces*. Others who have studied the theories and the theorists of evaluation credit her with influencing their thinking, such as Lee Cronbach in his 1982 book, *Designing Evaluations of Educational and Social Programs*, and with having brought a new perspective to bear on evaluation, that is, with being a Stage Two theorist—Will Shadish et al. in their 1991 book *Foundations of Program Evaluation*.

EP is indebted to Dr. Weiss for writing an "update" to the 1973 article, entitled "Politics and Evaluation: A Reprise with Mellow Overtones." Her 1992 reflections immediately follow the '73 work. We are also indebted to the Minneapolis Medical Research Foundation, Inc., for permission to publish the article which appeared in *Evaluation*, Vol. 1, 37–45. In turn, the article was based upon a paper that Dr. Weiss presented to the annual meeting of the American Psychological Association on August 30, 1973.

## Where Politics and Evaluation Research Meet

CAROL H. WEISS

Evaluation research is a rational enterprise. It examines the effects of policies and programs on their targets—whether individuals, groups, institutions, or communities—in terms of the goals they are meant to achieve. By objective and systematic methods, evaluation

---

Carol H. Weiss • Professor, Graduate School of Education, Harvard University, Cambridge, MA 02138.

*Evaluation Practice*, Vol. 14, No. 1, 1993, pp. 93–106.

ISSN: 0886-1633

Copyright © 1993 by JAI Press, Inc.

All rights of reproduction in any form reserved.

research assesses the extent to which goals are realized and looks at the factors that are associated with successful or unsuccessful outcomes. The assumption is that by providing "the facts," evaluation assists decision-makers to make wise choices among future courses of action. Careful and unbiased data on the consequences of programs should improve decision-making.

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

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

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

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

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

## PROGRAMS ARE POLITICAL CREATURES

Evaluation research assesses the effects of social programs, which in recent years have increasingly been governmental programs and larger in scale and scope than the programs studied in earlier decades. There have been important evaluations of job training programs, compensatory education, mental health centers, community health services, Head Start and Follow Through, community action, law enforcement, corrections, and other government interventions. Although there have been occasional studies of long-established traditional services, most evaluation efforts have been addressed to new programs; it is the program into which new money is being poured that tends to raise the most immediate questions about viability and continuation.

The programs with which the evaluator deals are not neutral, antiseptic, laboratory-type entities. They emerged from the rough and tumble of political support, opposition,

and bargaining. Attached to them are the reputations of legislative sponsors, the careers of administrators, the jobs of program staff, and the expectations of clients. The support of these groups coalesces around the program, but the counterpressures that were activated during its development remain active and the program remains vulnerable to interference from legislatures, bureaucracies, interest groups, professional guilds, and the media. It is affected as well by interagency and intraagency jockeying for advantage and influence.

The politics of program survival is an ancient and important art. Much of the literature on bureaucracy stresses the investment that individuals within an organization have in maintaining the organization's existence, influence, and empire. As Morton Halperin succinctly states:

Organizational interests, then, are for many participants a dominant factor in determining the face of the issue which they see and the stand which they take . . . Organizations with missions strive to maintain or to improve their (1) autonomy, (2) organizational morale, (3) organizational 'essence,' and (4) roles and missions. Organizations with high-cost capabilities are also concerned with maintaining or increasing (5) budgets.

It is not only around evaluation that social scientists bemoan the political factors that distort what they see as rational behavior. Economist Julius Margolis recently noted:

. . . you may go through a scientific analysis to answer the question of where the airport should be located, but an altogether different decision may finally emerge from the bureaucracy.

Bureaucrats, or in our terms program administrators and operators, are not irrational; they have a different model of rationality in mind. They are concerned not just with today's progress in achieving program goals, but with building long-term support for the program. This may require attention to factors and to people that can be helpful in later events and future contests. Administrators also have to build and maintain the organization—recruit staff with needed qualifications, train them to the appropriate functions, arrange effective interstaff relations and communications, keep people happy and working enthusiastically, and expand the influence and mission of the agency. There are budgetary interests, too, such as the need to maintain, increase or maximize appropriations for agency functioning. Clients have to be attracted, a favorable public image developed, and a complex system managed and operated. Accomplishing the goals for which the program was set up is not unimportant, but it is not the only, the largest, or usually the most immediate of the concerns on the administrator's docket.

Particularly when an organization is newly formed to run new programs, its viability may be uncertain. If the organization is dealing with marginal clientele, it can fall heir to the marginal repute of its clients, and it is likely to have relatively low public acceptance. Organizational vulnerability can become the dominant factor in determining what actions to take, and the need to build and maintain support can overwhelm the imperatives to achieve program goals.

In sum, social programs are the creatures of legislative politics and of bureaucratic politics. The model of the system that is most salient to program managers—and the components of the system with which they are concerned—are bound to be different from the model of the social scientist/evaluator. A program manager's view is probably no less rational. In fact, evidence suggests that programs can and do survive evaluations that show dismal failure to achieve goals. These programs, however, are less likely to survive a hostile congressional committee, newspaper exposes, or withdrawal of the support of professional groups.

There have been occasional references in evaluation literature to the need to recognize organizational "system" objectives as well as program goals (as in work by Herbert Schulberg and Frank Baker), but the notion has never caught on. So evaluators continue to regard these concerns of program staff as diversions from their true mission, and give them no points on the scorecard for effectiveness in the politics of organizational survival.

The disparity in viewpoint between evaluation researchers and program managers has consequences for the kind of study that is done, how well it is done, and the reception it gets when it is completed. Obviously the political sensitivities of program managers can dim their receptivity to any evaluation at all, and when a study is undertaken, can limit a program manager's cooperation on decisive issues of research design and data collection. Again, at the completion of the study, the program manager's political perspective will lessen the likelihood that he will view evaluative findings as conclusive or the need to act on them as imperative. Even rigorously documented evidence of outcomes may not outweigh all other interests and concerns.

More subtly, some political fallout shapes the very definition of an evaluation study. As an example, let us look at the specification of program goals that become the evaluator's criteria for effectiveness. Because of the political processes of persuasion and negotiation that are required to get a program enacted, inflated promises are made in the guise of program goals. Public housing will not just provide decent living space; it will improve health, enhance marital stability, reduce crime, and lead to improved school performance.

Furthermore, the goals often lack the clarity and intellectual coherence that evaluation criteria should have. Rather than being clear, specific, and measurable, they are diffuse and sometimes inherently incompatible. Again, it is the need to develop coalition support that leaves its mark. Holders of diverse values and different interests have to be won over, and in the process a host of realistic and unrealistic goal commitments are made.

Given the consequent grandiosity and diffuseness of program goals, there tends to be little agreement, even within the program, on which goals are real—real in the sense that effort is actually going into attaining them—and which are window-dressing. With this ambiguity, actors at different levels in the system perceive and interpret goals in different ways. What the Congress writes into legislation as program objectives is not necessarily what the Secretary's office or the director of the national program see as

their mission, nor what the state or local project managers or the operating staff actually try to accomplish.

The evaluator is faced with the task of sifting the real from the unreal, the important from the unimportant, perhaps even uncovering the covert goals that genuinely set the direction of the program (but are unlikely to surface in open discussion), and discovering priorities among goals. Unless he is astute enough to direct his research toward authentic goals, he winds up evaluating the program against meaningless criteria. Unless he is skillful enough to devise measures that provide valid indicators of success in this complex web of expectations, he runs the risk of having his report disowned and disregarded. It is not uncommon for evaluation reports to meet the disclaimer: "But that's not what we were trying to do."

While the evaluation study is in progress, political pressures can alter or undermine it. Let us look at one final example of how organizational politics can affect the shape of evaluation research. Programs do not always keep to their original course; over time, often a short span of time, they can shift in activities and in overall strategy and even in the objectives they seek to attain. They are responding to a host of factors: budget cutting or budget expansion, changes in administration or in top officials, veering of the ideological winds, changes in congressional support, public appraisal, initiation of rival agencies and rival programs, pervasive client dissatisfaction, or critical media coverage.

Whereas the evaluator wants to study the effects of a stable and specifiable stimulus, program managers have much less interest in the integrity of the study than in assuring that the program makes the best possible adaptation to conditions. This leaves the evaluator in a predicament: he is measuring outcomes of a "program" that has little coherence. What are the inputs? To what are the outcomes attributable? If the program succeeds, what activities should be replicated? If the program fails, what features were at fault? Unless programs under study are sheltered from the extremes of political turbulence, evaluation research produces outcome data that are almost impossible to interpret. On the other hand, to expect programs to remain unchanging laboratory treatments is to ignore the political imperatives. In this regard, as in others, programs have a logic and a rationality of their own.

## **THE POLITICS OF HIGHER ECHELON DECISION-MAKING**

Much evaluation research is sponsored not by individual projects or by managers of federal programs but by superordinate levels, such as the director of the agency or the Secretary or Assistant Secretary of the federal department, and the reports often go to cognizant officials in the Office of Management and Budget (OMB) and the White House and to members of congressional committees. If the organizations that run programs have a vested interest in their protection, these higher-level decision-makers can view the conclusions of evaluation research with a more open mind. They are likely to be less concerned with issues of organizational survival or expansion and more with

ensuring that public policies are worth their money and produce the desired effects. Of course, some legislators and Cabinet or sub-Cabinet officials are members of the alliance that supports particular programs. But it is generally true that the further removed the decision-maker is from direct responsibility for running the program, the more dispassionately he considers the evidence.

This of course does not mean that policy-makers venerate outcome data or regard it as the decisive input for decision. They are members of a policy-making system that has its own values and its own rules. Their model of the system, its boundaries and pivotal components, goes far beyond concern with program effectiveness. Their decisions are rooted in all the complexities of the democratic decision-making process: the allocation of power and authority, the development of coalitions, and the tradeoffs with interest groups, professional guilds, and salient publics. How well a program is doing may be less important than the position of the congressional committee chairman, the political clout of its supporters, or other demands on the budget. A considerable amount of ineffectiveness may be tolerated if a program fits well with prevailing values, if it satisfies voters, or if it pays off political debts.

What evaluation research can do is clarify what the political trade-offs involve. It should show how much is being given up to satisfy political demands and what kinds of program effects decision-makers are settling for-or foregoing when they adopt a position. It will not be the sole basis for a decision, and legitimately so. Other information and other values inevitably enter a democratic policy process. But evidence of effectiveness should be introduced to indicate the consequences that various decisions entail.

As a matter of record, relatively few evaluation studies have had a noticeable effect on the making and remaking of public policy. There are some striking exceptions, and in any case, our time frame may be too short. Perhaps it takes five or ten years or more before decision-makers respond to the accumulation of consistent evidence. There may need to be a sharp change in administration or a decisive shift in expectations. But to date, as Peter Rossi has pointed out, devastating evidence of program failure has left some policies and programs unscathed, and positive evidence has not shielded others from dissolution. Clearly, other factors weigh heavily in the politics of the decision process.

Perhaps one of the reasons that evaluations are so readily disregarded is that they address only official goals. If an evaluator also assessed a program's effectiveness in meeting political goals—such as showing that the Administration is “doing something,” or that the program is placating interest groups or enhancing the influence of a particular department—he might learn more about the measures of success that decision-makers value. He might learn why some programs survive despite abysmal outcomes, why some that look fine on indicators of goal achievement go down the drain, and which factors have the most influence on the making and persistence of policy. Just as economic cost-benefit analysis added the vital dimension of cost to analysis of outcomes, *political-benefit* analysis might help to resolve questions about political benefits and foregone opportunities.

It is true that many public officials in the Congress and the executive branch sincerely believe that policy choices should consistently be based on what works and what doesn't. It is also true that like all the other actors in the drama, policy-makers respond to the imperatives of their own institutions. One seemingly peripheral, but consequential factor, is the time horizon of the policy process. Presidents, governors, and legislators have a relatively short time perspective. They want to make a record before the next election. Appointed officials in the top positions of government agencies tend to serve for even shorter periods. The average tenure of officials in federal departments is a little over two years, as shown in a Brookings Institution study conducted by David Stanley, Dean Mann, and Jameson Doig. The emphasis therefore tends to be on take-offs, not on landings. It is often more important to a politically astute official to launch a program with great fanfare to show how much he is doing than to worry about how effectively the program serves people's needs. The annual cycle of the budget process also has the effect of foreshortening the time perspective. When decisions on funding levels have to be made within twelve months, there is little time to gather evidence (at least competent evidence) on program outcomes or to consider whatever information has been gathered.

What does it take to get the results of evaluation research a hearing? In a discussion of policy analysis (of which evaluation research is one phase), Charles Lindblom states that differences in values and value priorities constitute an inevitable limitation on the use of objective rational analysis. As I have already noted, maximizing program effectiveness is only one of many values that enter decisions. Therefore, Lindblom explains, the way that analysis is used is not as a substitute for politics but as a "tactic in the play of power":

It does not avoid fighting over policy; it is a method of fighting. . . . And it does not run afoul of disagreements on goals or values. . . because it accepts as generally valid the values of the policy-maker to whom it is addressed.

It does appear that evaluation research is most likely to affect decisions when the researcher accepts the values, assumptions, and objectives of the decision-maker. This means, obviously, that decision-makers heed and use results that come out the way they want them to. But it suggests more than the rationalization of predetermined positions. There is a further, important implication that those who value the *criteria* that evaluation research uses, those who are concerned with the achievement of official program goals, will pay attention as well. The key factor is that they accept the assumptions built into the study. Whether or not the outcome results agree with their own wishes, they are likely to give the evidence a hearing. But evaluation results are not likely to be persuasive to those for whom other values have higher priority. If a decision-maker thinks it is important for job trainees to get and hold on to skilled jobs, he will take negative evaluation findings seriously, but if he is satisfied that job training programs seem to keep the ghettos quiet, then job outcome data mean much less.

## THE POLITICS IMPLICIT IN EVALUATION RESEARCH

The third element of politics in the evaluation context is the stance of evaluation itself. Social scientists tend to see evaluation research, like all research, as objective, unbiased, and nonpolitical, as a corrective for the special pleading and selfish interests of program operators and policy-makers alike. Evaluation produces hard evidence of actual outcomes. But it incorporates as well a series of assumptions, and many researchers are unaware of the political nature of the assumptions they make and the role they play.

First, evaluation research asks the question: how effective is the program in meeting its goals? Thus, it accepts the desirability of achieving those goals. By testing the effectiveness of the program against the goal criteria, it not only accepts the rightness of the goals, it also tends to accept the premises underlying the program. There is an implicit assumption that this type of program strategy is a reasonable way to deal with the problem, that there is justification for the social diagnosis and prescription that the program represents. Further, evaluation research assumes that the program has a realistic chance of reaching the goals—or else the study would be a frittering away of time, energy, and talent.

For many programs, social science knowledge and theory would suggest that the goals are not well reasoned, that the problem diagnosis, the selection of the point of intervention, and the type of intervention are inappropriate, and that chances of success are slight. But when a social scientist agrees to evaluate a program, he gives an aura of legitimacy to the enterprise.

Furthermore, as Roland Warren has noted, the evaluator who limits his study to the effects of the experimental variables—those few factors that the program manipulates—conveys the message that other elements in the situation are either unimportant or that they are fixed and unchangeable. The intervention strategy is viewed as the key element, and all other conditions that may give rise to, sustain, or alter the problem are brushed aside. In particular, most evaluations—by accepting a program emphasis on services—tend to ignore the social and institutional structures within which the problems of the target groups are generated and sustained. Although evaluation studies can examine the effects of nonprogram variables, they generally concentrate on identifying changes in those persons who receive program services compared to those who do not, and they hold constant (by randomization or other techniques) critical structural variables in the lives of that particular population.

Warren suggests that there is an unhappy convergence between the preferred methodology of evaluation research—the controlled experiment—and the preferred method of operation of most single-focus agencies. Agencies tend to deal in piecemeal programs, addressing a single problem with limited intervention. He writes:

... for various reasons of practice and practicality they confine themselves to a very limited, relatively identifiable type of intervention, while other things in the life situation of the target population are . . . left unaltered. . . . The more piecemeal, the fewer the experimental variables involved, the more applicable is the [experimental] research design.



Methodologically, of course, experimental designs can be applied to highly complex programs (which is what factorial designs are about), but in practice there does seem to be an affinity between the experiment and the limited focus program. And if there is anything that we should have learned from the history of social reform, it is that fragmented program approaches make very little headway in solving serious social problems. An hour of counseling a week, or the introduction of paraprofessional aides, or citizen representation on the board of directors—efforts like these cannot possibly have significant consequences in alleviating major ills.

Another political statement is implicit in the selection of some programs to undergo evaluation, while others go unexamined. The unanalyzed program is safe and undisturbed, while the evaluated program is subjected to scrutiny. What criteria are used in selecting programs to evaluate? Obviously, it is the new and (perhaps) innovative program that is put on trial while the hardy perennials go on, whether or not they are accomplishing their goals, through the sheer weight of tradition.

Other criteria for selecting programs for evaluations are even more overtly political. Thus in a discussion of program analysis, Charles Schultze makes two recommendations: (1) program analysts should give more consideration to programs that do not directly affect the structure of institutional and political power than to programs that fundamentally affect income distribution or impinge on the power structure, and (2) analysts can be more useful by studying new and expanding programs than long existing programs with well-organized constituencies.

There are persuasive reasons for such prescriptions. Evaluators, like all other analysts who ignore the political constraints of special interests, institutional power, and protective layers of alliances may confront the decision-maker with troublesome information. If time after time they bring in news that calls for difficult political choices, if they too often put the decision-maker in a position that is politically unviable, evaluators may discredit evaluation research as a useful tool. Nevertheless, there are serious political implications in restricting evaluation to the unprotected program and the program that is marginal to the distribution of economic and political power.

The structure of the evaluation research enterprise also has political overtones. To begin with, evaluation is generally commissioned by the agency responsible for the program, not by the recipients of its efforts. This is so obvious and taken for granted that its implications are easily overlooked. Some of its consequences, however, are that the officials' goal statements form the basis for study and if recipients have different needs or different ends in mind, these do not surface. Another probability is that the evaluator interprets his data in light of the contingencies open to the agency. The agency is the client and the evaluator tries to gear his recommendations to accord with realistic practicalities. Furthermore, he reports study findings to decision-makers and managers, usually not to program participants; if the findings are negative, officials may not completely bury the report (although sometimes they try), but they can at least release it with their own interpretations: "We need more money," "We need more time," or "The evaluation was too crude to measure the important changes that took place."

To the extent that administrators' interpretations shape the understanding of a study's import, they constrain the decisions likely to be made about that program in the future and even to influence the demands of the target groups. An evaluation report showing that Program A is doing little good, if interpreted from the perspective of the participants in the program, might well lead to very different recommendations from those developed by an agency-oriented evaluator or a program official.

Most of these political implications of evaluation research have an "establishment" orientation. They accept the world as it is: as it is defined in agency structure, in official diagnoses of social problems, and in the types of ameliorative activities that are run. But the basic proclivity of evaluation research is reformist. Its whole thrust is to improve the way that society copes with social problems. At the same time that evaluation research accepts program assumptions, it also subjects them to scrutiny; its aim is to locate discrepancies between intent and actual outcome.

In addition to this reformist thrust, Harold Orlans has indicated that social science evaluators tend to be more liberal in orientation than many of the agencies they study. And their perspectives inevitably affect their research. As social scientists increasingly recognize, no study collects neutral "facts": all research entails value decisions and to some degree reflects the researcher's selections, assumptions, and interpretations. This liberal bias of much evaluation research can threaten its credibility to officialdom. Thus, Laurence Lynn, Jr., a federal Assistant Secretary writes:

The choices of conceptual frameworks, assumptions, output measures, variables, hypotheses, and data provide wide latitude for judgment, and values of the researcher often guide the decisions to at least some degree. Evaluation is much more of an art than a science, and the artist's soul may be as influential as his mind. To the extent that this is true, the evaluator becomes another special interest or advocate rather than a purveyor of objectively developed evidence and insights, and *the credibility of his work can be challenged*. [Italics added]

In this statement, there seems to be an assumption that such a thing as "objectively developed evidence" exists and that assumptions and values are foreign intrusions. But the message that comes through is that "objectively developed evidence" is that which develops only out of government-sanctioned assumptions and values. Certainly evaluators funded by government have an obligation to start with the official framework, but they should be able to look at other variables and other outcomes, wanted and unwanted, in addition to those set by official policy.

The intrinsically reformist orientation of evaluation research is apparent in its product. Evaluation conclusions are the identification of some greater or lesser shortfall between goals and outcomes, and the usual recommendations will call for modifications in program operation. The assumptions here are (1) that reforms in current policies and programs will serve to improve government performance without drastic restructuring and (2) that decision-makers will heed the evidence and respond by improving programming. It is worthwhile examining both these assumptions, particularly when we take note of one major piece of intelligence: evaluation research discloses that most programs dealing with social problems fail to accomplish their goals. The finding of little impact is pervasive over a wide band of program fields and program strategies.

True, much of the evaluation research has been methodologically deficient and needs upgrading. (There is an extensive literature on methodological shortcomings. Donald Campbell and Selma Mushkin are among those who have written cogent critiques.) But there is little evidence that methodologically sounder studies find more positive outcomes. Numbers of excellent studies have been carried out, and they generally report findings at least as negative as do the poor ones. Moreover, the pattern of null results is dolefully consistent. So despite the conceptual and methodological shortcomings of many of the studies, the cumulative evidence has to be taken seriously.

What does the evaluation researcher recommend when he finds that the program is ineffective? For a time, it may be a reasonable response to call attention to possible variations that may increase success—higher levels of funding, more skilled management, better trained staff, better coordination with other services, more intensive treatment, and so on. If these recommendations are ignored, if the political response is to persist with the same low cost, low-trouble program, there is not much more that the social scientist can learn by evaluating participant outcomes. If program changes are made, then further evaluation research is in order. But there comes a time when scores or even hundreds of variants of a program have been run, for example, in compensatory education or rehabilitation of criminal offenders, and none of them has shown much success. If it was not evident before, it should be clear by then that tinkering with the same approaches in different combination is unlikely to pay off.

There needs to be serious reexamination of the basic problem, how it is defined, what social phenomena nurture and sustain it, how it is related to other social conditions and social processes, and the total configuration of forces that have overwhelmed past program efforts. Fragmented, one-service-at-a-time programs, dissociated from people's total patterns of living, may have to be abandoned, and as Daniel Moynihan has suggested, integrated policies that reach deeper into the social fabric will have to be developed. What this suggests is that in fields where the whole array of past program approaches has proved bankrupt, the assumption is no longer tenable that evaluation research of one program at a time can draw useful implications for action or that piecemeal modifications will improve effectiveness.

As for the other major premise on which the utility of evaluation research is based—that policy-makers will heed research results and respond by improving programming—there is not much positive evidence either. I have noted how the politics of program survival and the politics of higher policy-making accord evaluative evidence relatively minor weight in the decisional calculus. It is when evaluation results confirm what decision-makers already believe or disclose what they are predisposed to accept that evaluation is most apt to get serious attention. Thus, for example, the Nixon Administration was willing to listen to the negative findings about the Johnson Great Society programs. As Allen Schick has noted, evaluation research is comfortably compatible with a government perspective of disillusionment with major program initiatives—with stock-taking and retrenchment. As a consequence, the fiscal year 1973 budget submitted to Congress proposed to cut out or cut back programs that

weren't working. The evaluation researcher—now that somebody was paying attention to findings—was cast in the role of political hatchet man.

Because evaluation researchers tend to be liberal, reformist, humanitarian, and advocates of the underdog, it is exceedingly uncomfortable to have evaluation findings used to justify an end to spending on domestic social programs. On the other hand, it is extremely difficult for evaluators to advocate continuation of programs that they have found had no apparent results. The political dilemma is real and painful. It has led some social scientists to justify continued spending on avowedly ineffective programs to preserve the illusion that something is being done. Others have called for continued spending, whatever the outcome, so as not to lose the momentum of social progress. Others justify the programs with explanations that they regarded as specious when used by program staff: the programs serve other purposes, the evaluations aren't very good, the programs need more money, they need more time. My own bent is to find some truth in each of these justifications, but they tend to be declarations based on social ideology and faith. Evaluators can maintain them only so long without providing evidence that these factors are responsible for the poor showing or that the programs are achieving other valued ends.

What would be a responsible position for evaluation research? It seems to me that there are a few steps that can be taken. One reform in evaluation research would be to put program goals in sensible perspective. Among the many reasons for the negative pall of evaluation results is that studies have accepted bloated promises and political rhetoric as authentic program goals. Whatever eager sponsors may say, day care centers will not end welfare dependency and neighborhood government will not create widespread feelings of citizen efficacy. Programs should have more modest expectations (helping people to cope is not an unimportant contribution), and they should be evaluated against more reasonable goals. Furthermore, evaluations that average the effects of numbers of local projects and come up with summary "pass/fail" measures are not likely to be optimally useful. More learning will come from specifying the conditions that are associated with better or poorer outcomes—conditions of program operation as well as conditions in the larger social context.

A further step along this course would be to evaluate a particularly strong version of the program before, or along with, the evaluation of the ordinary levels at which it functions. This would tend to show whether the program *at its best* can achieve the desired results, whether accomplishments diminish as resource level or skills decline, and how intensive an effort it takes for a program to work. If the full-strength "model" program has little effect, then it is fruitless to tinker with modest, low-budget versions of it.

More fundamentally, however, it seems to me that now in some fields there is a limit to how much more evaluation research can accomplish. In areas where numbers of good studies have been done and have found negative results, there seems little point in devoting significant effort to evaluations of minor program variants. Evaluation research is not likely to tell much more. There is apparently something wrong with many of our social policies and much social programming. We do not know *how* to solve some of the major problems facing the society. Nor do we apply the knowledge

that we have. We mount limited-focus programs to cope with broad-gauge problems. We devote limited resources to long-standing and stubborn problems. Above all, we concentrate attention on changing the attitudes and behavior of target groups without concomitant attention to the institutional structures and social arrangements that tend to keep them "target groups."

For the social scientist who wants to contribute to the improvement of social programming, there may be more effective routes at this point than through evaluation research. There may be greater potential in doing research on the processes that give rise to social problems, the institutional structures that contribute to their origin and persistence, the social arrangements that overwhelm efforts to eradicate them, and the points at which they are vulnerable to societal intervention. Pivotal contributions are needed in understanding the dynamics of such processes and in applying the knowledge, theory, and experience that exist to the formulation of policy. I suspect that in many areas, this effort will lead us to think in new categories and suggest different orders of intervention. As we gain deeper awareness of the complexities and interrelationships that maintain problem behavior, perhaps we can develop coherent, integrated, mutually supportive sets of activities, incentives, regulations, and rewards that represent a concerted attack and begin to deserve the title of "policy."

How receptive will established institutions be to new ways of looking at problems and to the new courses of action that derive from them? I suggested earlier that decision-makers tend to use research only when its results match their preconceptions and its assumptions accord with their values. There will certainly be resistance to analysis that suggests changes in power relations and in institutional policy and practice. But legislatures and agencies are not monoliths and there may well be some supporters, too. As time goes on, if confirming evidence piles up year after year on the failures of old approaches and if mounting data suggest new modes of intervention, this will percolate through the concerned publics. When the political climate veers toward the search for new initiatives, or if sudden crises arise and there is a scramble for effective policy mechanisms, some empirically grounded guidelines will be available.

Of course, there remains a vital role for evaluation research. It is important to focus attention on the consequences of programs, old and new, to keep uncovering their shortcomings so that the message gets through, and to locate those programs that do have positive effects and can be extended and expanded.

It is important to improve the craft of evaluation so that we have greater confidence in its results. To have immediate and direct influence on decisions, there is a vital place for "inside evaluation" that is consonant with decision-makers' goals and values—and perhaps stretches their sights a bit. There is also a place for independent evaluation based on different assumptions with wider perspectives, and for the structures to sustain it. One of the more interesting roles for evaluation is as "social experimentation" on proposed new program ventures—to test controlled small-scale prototypes before major programs are launched and thereby gain good measures of their consequences.

Nevertheless, given the record of largely ineffective social programming, I think the time has come to put more of our research talents into even earlier phases of the

policy process, into work that contributes to the development of ideas and prototypes. I believe that we need more research on the social processes and institutional structures that sustain the problems of society and closer social science involvement in the application of that research. I have hope that this can contribute to understanding which factors have to be altered if change is to occur and, in time, to more effective program and policy formation.

## REFERENCES

- Campbell, D.T. and Erlebacher, A. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In: J. Hellmuth (Ed.), *Compensatory Education: A National Debate*, Vol. 3, New York: Brunner/Mazel.
- Caplan, N. and Nelson, S. (1973). On being useful: The nature and consequences of psychological research on social problems. *American Psychologist*, 28:(3), 199–211.
- Halperin, M.H. (1971). Why bureaucrats play games. Reprint 199. Washington, D.C. Brookings Institution, 1971.
- Hatry, H.P., Winnie, R.E. and Fisk, D.M. (1973). *Practical program evaluation for state and local government officials*. Washington: The Urban Institute.
- Lindblom, C.E. (1968). *The policy-making process*. Englewood Cliffs, N.J.: Prentice-Hall.
- Lynn, L.E. Jr. (1973). A federal evaluation office? *Evaluation* 1(2), 56–59, 92, 96.
- Margolis, J. (1971). Evaluation criteria in social policy. In: T.R. Dye (Ed.), *The Measurement of Policy Impact*. Tallahassee: Florida State University, 25–31.
- Moynihán, D.P. (1970) Policy vs. program in the 70's. *The Public Interest*. (No. 20, 1970), 90–100.
- Mushkin, S.J. (1973). Evaluations: Use with caution. *Evaluation* 1(2), 30–35.
- Orlans, H. (1973). *Contracting for knowledge*. San Francisco: Jossey-Bass.
- Rossi, P. (1969). Practice, method, and theory in evaluating social-acting programs. In: J.L. Sundquist (Ed.), *On Fighting Poverty: Perspectives from Experience*. New York: Basic Books, 217–34.
- Schick, A. (1971) From Analysis to Evaluation. *Annals of the American Academy of Political and Social Science* (Vol. 394) 57–71.
- Schulberg, H.C. and Baker, F. (1968) Program Evaluation Models and the Implementation of Research Findings. *American Journal of Public Health* (58:7) 1248–55.
- Schultze, C.L. (1968) *The politics and economics of public spending*. Washington: Brookings Institution.
- Stanley, D.T., Nann, D.E., and Doig, J.W. (1967) *Men who govern: A biographical profile of federal political executives*. Washington, Brookings Institution.
- Ward, D.A. and Kassebaum, G.G. (1972). On biting the hand that feeds: Some implications of sociological evaluations of correctional effectiveness. In: C. Weiss (Ed.), *Evaluating action programs: Readings in social action and education*. Boston: Allyn and Bacon, 300–10.
- Warren, R. (1973). The social context of program evaluation research. Paper presented at Ohio State University Symposium on Evaluation in Human Service Programs, June.
- Weiss, C.H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, N.J.: Prentice-Hall.
- Weiss, C.H. (1970). The politicization of evaluation research. *Journal of Social Issues* 26(4), 57–58.
- Weiss, C.H. (1973). The Politics of Impact Measurement. *Policy Studies Journal*, 1(3), 179–83.