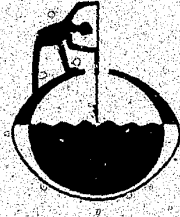
ISEGR OCCASIONAL PAPERS

No. 6, February 1972

The Problem of Measuring the Impacts of Social-Action Programs

THOMAS A. MOREHOUSE



NSTITUTE OF SOCIAL, ECONOMIC
AND GOVERNMENT RESEARCH
UNIVERSITY OF ALASKA
Fairbanks, Alaska

ISEGR Occasional Papers are published periodically by the Institute of Social, Economic and Government Research, University of Alaska. Authors are free to develop their own ideas on their own topics.

Thomas A. Morehouse is an associate professor of political science at the Institute of Social, Economic and Government Research. He holds a Ph.D. and a M.A.P.A. from the University of Minnesota and a B.A. from Harvard College.

Preparation of this paper was supported in part by the U.S. Department of Health, Education and Welfare under Title I of the Higher Education Act of 1965. The paper should not be construed as representing the opinions or policy of any government agency.

Price $1.00

Victor Fischer, Director of the Institute

James D. Babb, Jr., Editor

2/24/25

THE PROBLEM OF MEASURING THE IMPACTS OF SOCIAL-ACTION PROGRAMS

by

Thomas A. Morehouse

Acknowledgements

The author wishes to thank Gordon Harrison, Arthur Hippler, Judith Kleinfeld, and Arlon Tussing for their helpful comments on earlier drafts of this paper.

INTRODUCTION

Since the passage of the Economic Opportunity Act of 1964, which established the Office of Economic Opportunity (OEO) and began the "war on poverty," Alaska and the rest of the nation have witnessed a proliferation of innovative social action and economic development programs. These have included manpower training and job development, pre-school education and child care, community organizational activity, legal services, and attempts to revitalize depressed local and regional economies, to strengthen public educational programs, to increase the supply of housing for the poor, and a host of others.

Most of these programs have not had as their primary purpose the delivery of tangible goods and services to needy populations. Instead, they have attempted to provide the means and opportunities for people eventually to obtain such goods and services for themselves. The concern of this paper, however, is not the number and kinds of attempts underway in Alaska and throughout the nation to deal with social and economic problems. Rather, it is concerned with the question of determining how adequate and effective such efforts may be. In short, how are the accomplishments of these programs measured?

"Program evaluation research" has emerged as the means of dealing with the question of program effectiveness. Its purpose is to determine what works and how well, and what does not work. More specifically, evaluation research is defined as the means of determining the extent to which a program is achieving its goals, using "methods that yield evidence that is objective, systematic, and comprehensive."[1] It is intended to focus on the "results" of program efforts, assessing them in relation to program objectives.

---

[1]Francis C. Caro, "Approaches to Evaluation Research: A Review," *Human Organization*, Vol. 28 (Summer 1969), p. 87.

One critical assumption of evaluation research -- when defined as the measurement of results -- is that a program has a specific objective or objectives. A second critical assumption is that the intended or predictable results will be substantial enough to warrant attempts to locate and measure them. Evaluation of many of the newer social action and economic development programs in recent years indicates that, all to often, neither assumption holds up. In other words, program aims have often been too broad or ambiguous, and "results" too weak or elusive to justify the kinds of complex and costly evaluation efforts that have been made.

This paper provides guidance for program administrators in determining when evaluation research is needed, what it can do, what it cannot do, and what alternative forms it can take. As new social and economic programs are initiated in Alaska during the 1970's, the need for research on program effectiveness will increase accordingly. But in view of the very uneven record of such research to date, so also will the need to avoid wasteful or misdirected research and resulting frustration on the part of both administrators and researchers.

THE EMERGENCE OF PROGRAM EVALUATION

During the late 1960's, federal laws and regulations increasingly required that federal domestic programs be evaluated to determine how effectively they were accomplishing their objectives. Beginning with amendments to the poverty program legislation in 1967, evaluation requirements have been incorporated in a wide range of programs, particularly the more innovative social action and economic development programs. In addition, several programs initially intended as full-scale efforts to eradicate poverty, renovate slum neighborhoods, and revitalize local or regional economies were later redefined as less costly and more limited "pilot," "research," or "demonstration" programs. Regarded as "social experiments," the programs had to be carefully evaluated to determine their effectiveness before substantial resources could be committed to them.

Most evaluation research on federal programs has failed to provide the clear-cut assessments of program effectiveness

it was supposed to deliver. And, not only has most of the research work been considered unsuccessful, but even those instances of technically "successful" evaluation research have failed to affect program policy making or administration in any significant way. Evaluation work can therefore "fail" in two ways: it can, for methodological reasons, fail to identify and measure relevant effects of programs; and it can fail to produce "acceptable" findings consistent with the policy makers' commitments to program success.[2]

In many cases, both of these reasons for "failure" may stem from the fact that evaluation research has been applied to programs whose planned or intended effects are slight and therefore hard to locate and measure.[3] A National Science Foundation study group found that even "when carefully designed research studies have been carried out, they have often shown the programs have been only minimally effective."[4] And a comprehensive study of federal evaluation

---

[2]See, among others, National Science Foundation, *The Nation's Use of the Social Sciences*, Report of the Special Commission on the Social Sciences of the National Science Board (Washington: U.S. Government Printing Office, 1969); Joseph S. Wholey, and others, *Federal Evaluation Policy: Analyzing the Effects of Public Programs* (Washington: The Urban Institute, 1970); Marshall S. Smith, and Joan S. Bissell, "Report Analysis: The Impact of Head Start," *Harvard Educational Review*, Vol. 40, No. 1 (February 1970), pp. 51-104; Walter Williams and John Evans, "The Politics of Evaluation: The Case of Head Start," *The Annals*, Vol. 385 (September 1969), pp. 118-132; and Reginald K. Carter, "Clients' Resistance to Negative Findings and the Latent Conservative Function of Evaluation Studies," *The American Sociologist*, Vol. 6 (May 1971), pp. 118-124.

[3]This is not to say that the community action or model cities programs, for example, do not generate a great deal of local activity. Indeed, the community action program has been quite visibly active, in many places even to the point of openly challenging city hall and other "establishment institutions." The point here is that the *planned* or *intended* effects of the programs have been reported to be minimal.

[4]National Science Foundation, *op. cit.*, p. 76.

policy concluded that "the experience of social program evaluations in the past few years would seem to indicate that, *a priori*, one cannot expect much in the way of highly valued benefits as the results of a social program."[5]

There may be some circularity here, since the judgments of program "impact" or "benefits" are based on evaluation work that may itself be unsuccessful in identifying and measuring results considered most relevant by administrators and others involved with the program. Moreover, programs often have significant effects that are neither clearly intended nor easily measurable.[6] The result often is that an imperfect tool (evaluation) is brought to bear on programs whose planned or intended objectives, means, and impacts may be diffuse, complex, or minimal, or all of these.

Where, then, does the fault lie -- with the programs, with the evaluation research, or with both? An adequate answer requires consideration both of research methodology and of the programs to which it is applied.

THE LIMITS OF SCIENTIFIC EVALUATION

Social scientists are inclined to distinguish between two basic methodological orientations. One emphasizes experimentation and quantitative measures and may be referred to as "quasi-experimental." Although based on experimental models, it involves important modifications or compromises of classical experimental designs derived from the natural sciences. The second consists of many varying approaches, all of which share in common the absence of experimentation and quantitative measurement as central characteristics of research strategy. Experimentation and quantitative measurement have been the most widely sanctioned means of conducting evaluation research-- notwithstanding that performance has often fallen

[5]Wholey, and others, *op. cit.*, p. 91.

[6]Cf. Donald T. Campbell, "Considering the Case Against Experimental Evaluations of Social Innovations," *Administrative Science Quarterly*, Vol. 15 (March 1970), p.111.

far short of aspirations and that many evaluation research projects have been carried out in other, less highly regarded ways.[7]

The experimental approach to evaluation assumes that a program has well defined objectives and that its effectiveness can be determined by measuring the extent to which the objectives are achieved. For this approach to be successful, a program should meet at least four criteria. First, and most important, there should be a clear and precise statement of the specific results intended by the program. Obviously, if the principal purpose of evaluation is to determine the extent to which a program achieves its intended objectives, then objectives must be clearly defined and stated in such a way as to permit the necessary measurements to be made.

Second, the situation or setting of the program should be reasonably "controlled," i.e., not subject to unknown idiosyncratic forces that could disrupt, obscure, or otherwise significantly influence program effects. Related to this, another group and setting, similar to the program group setting, may be identified and studied concurrently. The comparison of effects in an "experimental" group (one subject to a program "treatment") with those in a "control" group (one not exposed to the "treatment") may be the best way to explore whether the changes that occur in the program situation are actually the result of the program and not of outside forces or influences. For example, was it primarily a manpower training program or changes in the general economy that produced an increase in the employment of persons served by the program?

Third, the program "treatment" should be reasonably uniform. If the treatment varies in unplanned or unintended ways for different subjects at different times, then it may

[7]Caro, *op. cit.*, provides a comprehensive review of the current evaluation research literature, discussing the pros and cons, the problems and pitfalls, of the experimental and quasi- or non-experimental approaches. Consistent with most of the authorities he cites, Caro concludes somewhat vaguely that "scientific" evaluation research may not reach methodological perfection, but it is nonetheless to be preferred to "impressionistic accounts." Caro does not define what is meant by either of the terms in quotes, or if, indeed, these are meaningful alternatives.

become impossible to say just what has caused what. In other words, if a program is to be evaluated, it must have some determinate, stable form to assure that evaluators and administrators are talking about the same thing.[8]

Fourth, the program treatment should be applied in a large enough number of cases to provide an adequate sample of program experience. If the sample is too small, it may be impossible to tell whether program effects were representative, and evaluators may fall into the familiar problem of trying to draw general conclusions from a single or small number of cases.

Ideally, then, an experimental approach would be used to evaluate a program that has a distinct objective and standardized treatment, operates in controlled settings, and includes enough cases for valid generalization.[9] Most of the federal social action and economic development programs to which evaluation requirements have been attached do not fit this pattern.[10] Instead, they often have broad aims that are not, and — without distortion or change — perhaps cannot be, specified in clear-cut form. Such programs may be concerned

[8]A program might deliberately and systematically take different forms in different places in order that administrators can compare the relative effectiveness of different strategies or components. If this is the case, it should immediately be apparent that the research problem of control-comparisions is compounded.

[9]Robert S. Weiss and Martin Rein, "The Evaluation of Broad-Aim Programs: A Cautionary Case and a Moral," *The Annals*, Vol. 385, (September 1969), p. 134.

[10]Although rare, such programs, carefully designed as experiments, do exist. A prominent example is the "negative income tax" program conducted in New Jersey by OEO. The evaluation research focuses on a single factor, the effect of negative taxation on the incentive to work. The treatment -- the provision of money to individuals -- is clearly standardized, and an adequate number of cases can be studied in reasonably controlled situations. See Robert S. Weiss and Martin Rein, "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties, and an Alternative," *Administrative Science Quarterly*, Vol. 15 (March 1970), pp. 104-105.

more with changing many aspects of a general social and economic setting than with affecting individuals directly. They generally operate in open, unique, and uncontrolled community and regional settings, and tend to take different forms in different situations. Finally, the very fact that a program is present and operating in a given setting may be *the* factor, above all others, that makes the situation incomparable.

Thus, there have been a large number of unsuccessful evaluation projects, in part because the methods have been a poor match for the complex and variable program situations to which they have been applied. Evaluators and their patrons seem to have assumed that there are no essential differences between programs with clear-cut aims and concrete, measurable effects and programs with broad aims and elusive effects. The programs of the 1960's, on which evaluation efforts have focused, are mostly of the latter kind. This record of evaluation research failure, in turn, may have been perpetuated because so little has been done to develop more flexible and imaginative approaches to evaluation that researchers, administrators, and policy makers might accept as legitimate.

OPPORTUNITY AND MAINTENANCE PROGRAMS

Inappropriate methodology may thus be one source of evaluation research failures. But, as indicated, it is difficult to separate the methodological problem from the "program problem." If the program is indeed too weak (insufficient and/or misdirected resources) to have significant planned effects, then no research directed specifically to measuring intended effects can make the program appear successful in meeting its objectives. The researcher, in reporting that the program has had little or no effect, thus risks the ire of program administrators and further attacks on the adequacy of his methodology.[11]

[11]Some programs, of course, have broader aims and more elusive results than do others. Examples of the former type would be OEO's community action program, the Department of Housing and Urban Development's model cities planning program, the

The specificity or breadth -- the scope -- of program aims is, as discussed above, one factor to be considered in determining the feasibility and kind of evaluation research worth doing. Closely related to, but distinguishable from this factor, is the nature of program aims or objectives. Social and economic programs may be identified either as maintenance programs or as opportunity programs.[12]

A maintenance program is one that provides tangible goods or services to a clearly defined population. Its "primary goal is to deliver a service (or good) that is itself a highly valued commodity, for example, money or food."[13] An opportunity program, in contrast, is concerned not with directly meeting the material needs of a group, but with increasing the group's capabilities or opportunities to bring about a "positive change in an individual's capacity to earn or to learn."[14] While a maintenance program may, for example, distribute surplus food commodities or provide housing to a needy population, an opportunity program may seek to equip a person to obtain employment through job training, adult education, or other means, so that the individual may earn the income necessary to buy adequate food or housing.

The most significant issues and difficult problems of evaluation arise in connection with programs defined in opportunity terms. Because most of the innovative anti-poverty and other programs of the 1960's attempted to provide "opportunities" for people eventually to obtain goods and services for themselves, the programs' intended "outputs" have often been elusive and difficult to sort out and measure. And, regardless of the maintenance effect (which, as a practical matter, may be the primary effect) of many programs, federal agencies have insisted that

Economic Development Administration's district planning program, and OEO's community enterprise development corporation program. Examples of relatively more focused programs are manpower training, pre-school education, rural housing, and legal services.

[12]Williams and Evans, *loc cit.*

[13]*Ibid.*, p. 131.

[14]*Ibid.*

opportunity purposes are paramount and that evaluations should be designed accordingly. Yet opportunity and maintenance activities are not mutually exclusive; both may be directed to a given group at the same time, and they may be parts or aspects of the same program. Evaluators consequently find themselves groping for new ways of evaluating such programs, or attempting to use a conventional quasi-experimental research design where it will not work.

In Alaska, for example, the federally supported Alaska Village Electric Cooperative (AVEC) is constructing community electric power plants in some 50 remote Native villages of from 200 to 400 people. Defined as a "maintenance" program, AVEC is providing a basic service, subsidized by the government, and the service is valued directly for its own sake. On the other hand, OEO policy [15] emphasizes "opportunity" aspects of the program, in accordance with the rationale for most of the federal social action and economic development programs of the 1960's. OEO states that investment in facilities "simply for the purpose of easing the life of isolated poor Indians is difficult if not impossible for government agencies to justify."[16] There must be an economic growth rationale. Thus, the AVEC program "is testing whether there are economic activities which are presently not feasible, but which can be made feasible by the provision of electric power."[17] In addition, OEO has promoted certain community organizational objectives of the program, so that "social development" is added to "economic development" as opportunity purposes. Whether to define a program like AVEC either in "maintenance" terms or in "opportunity" terms, or both, is a very critical step in determining the form and character (and the feasiblity) of evaluation. The point is that how a program is defined goes far to determine what approach should be taken to the problem of evaluation research.

[15]As set forth in U.S. Office of Economic Opportunity, "Highlight Memorandum," Washington, D.C., April 23, 1968.

[16]*Ibid*

[17]*Ibid*

If it is agreed that definition of program objectives is the critical factor in evaluation research, and that the objectives themselves raise important research questions, it follows that program evaluation may need to focus as much on objectives as results. The researcher must determine what it is that program policy makers and administrators expect to happen as results of the program, why the results are desired, how the results are to be obtained, and when they are expected to occur. And, in the course of his research, the evaluator must regard the intended effects of the program as but one indicator of program performance. Just as important may be the program's *unintended* effects, and the *process* that relates objectives and means to one another and that fits them to problems in varying environments.

Many programs are not finished products ready to be "tested" by measuring specified effects in terms of clear-cut objectives. Instead, they are explorations of problems, objectives, and means. They are, in this sense, at least as much process-oriented as they are results-oriented.[18] Thus, a more broadly conceived approach to evaluation research may well be essential at this time, since so many recent programs are efforts to develop new administrative processes directed to newly defined problems, rather than full-scale commitments of resources designed to achieve maximum impacts on clearly targeted problem areas and groups.

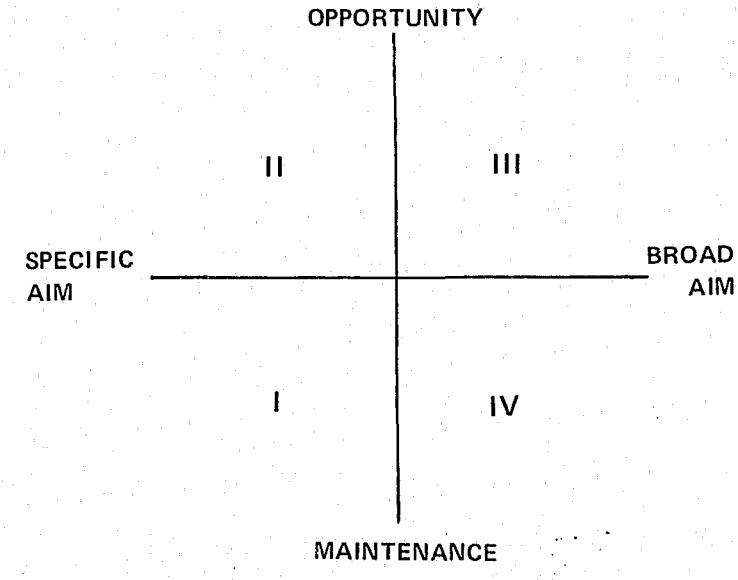
TYPES OF PROGRAMS

The definition of programs in terms of the *scope* of program aims -- broad or specific -- and the *nature* of program purposes -- opportunity or maintenance -- can be of considerable assistance in determining the purposes, feasibility, and types of evaluation or other program research that should be conducted. These definitions lead to four basic types of program categories as depicted in the following diagram:

[18]For an excellent discussion of social-action programs as "explorations," see Peter Marris and Martin Rein, *Dilemmas of Social Reform* (New York: Atherton Press, 1969), pp. 203-207.

FIGURE 1

Types of Programs



In reality, highly variable programs do not fall neatly into a small set of separate boxes. The four categories represent matters of "more or less" emphasis, rather than discrete cases. Their analytical purpose is to help identify the critical problems of evaluation research that are associated with the major program types.

Type I: Specific Aim — Maintenance Programs

Specific aim—maintenance programs include such things as the provision of rent supplements, legal services, food stamps, surplus food distribution, housing, preventive medicine, and community facilities, such as water and sewer services. Because such programs usually deliver tangible goods or services to a definite population, their aims or objectives often can be precisely specified. Thus, Type I programs are generally the easiest to plan, organize, administer, and evaluate. In fact, after an initial shakedown period, the

administrative system for the program should itself provide directly for the continuing data collection and analysis needed to determine how effectively the program is achieving its aims.

However, this is not to say that at no point would a specially organized evaluation project be desirable and useful, and probably best carried out by an outside evaluator or team. For example, whether a food stamp program is actually reaching those most in need may be a very real and critical issue, and one that program administrators may not themselves have the time, other resources, or inclination to investigate. Further, it is a common occurrence that initially high levels of administrative enthusiasm wear down. Administration becomes routine, yet changes occur in the character and extent of the problem and in the population to which the program is directed. These and other factors indicate a need for periodic program assessment, of which "evaluation" — strictly defined as measurement of "outputs" — may not be the most prominent part.

Type II: Specific Aim — Opportunity Programs

It is with Type II programs that the most significant problems of evaluation (as well as of program planning and administration) begin to appear. Specific aim—opportunity programs do not have a tangible product or output that is valued primarily for its own sake. Instead, the intent of the program is to enhance the opportunities of individuals, groups, communities, or regions to achieve some further goal and thus better their socio-economic or political condition.

Examples of specific aim—opportunity programs are manpower training to equip individuals to obtain jobs and earn income; community facilities programs intended to provide the infrastructure or support facilities (e.g. power plant) that will lead to new employment and income producing opportunities in a community or region; and pre-school education programs, such as Head Start, that attempt to prepare a child to respond more effectively to school. In contrast to Type I programs that directly provide goods and services immediately beneficial in themselves, Type II programs are, in effect, linkages between an individual or group and a further, specified goal.

To the extent that the program does indeed have a "specific aim," one prerequisite of effective program evaluation can be met: a clear statement of program objectives. However, official statements are often not as definitive as they may first appear. For example, the stated objective of a manpower training program may be to train and place in jobs a given number of persons. But there are other and more complex questions to answer than merely how many people have been trained and how many have actually been placed in jobs. Knowledge of the characteristics of trainees (e.g., previous employment experience), types of jobs in which trainees have been placed, length of time between job training and placement, levels of salaries or wages, length of time on the job, and job satisfaction are also likely to be very significant in assessing program performance. To judge a program solely in terms of number of placements may give a false impression of program success.

This example suggests a further problem for administrators and evaluators of most opportunity programs. When are the results expected to occur? Can or should program effects be further classified as immediate, intermediate, and long-term? What time period should be used? What relative weights should be assigned to effects occurring over an extended period of time?

Type III: Broad Aim — Opportunity Programs

The problems of evaluating specific aim—opportunity programs are present in the case of broad aim—opportunity programs as well, only more so. The Department of Housing and Urban Development's Model Cities Program is a good Type III example. It is primarily a broad aim—opportunity program because it attempts to reform a multi-faceted urban planning process, which, in turn, is intended to achieve improved coordination among city, state, and federal agencies; citizen participation; increased agency responsiveness to local needs; increased federal support; and other goals. This program does have tangible components such as urban renewal, housing, community facilities, and the like, but they are not provided separately or for their own sake. One of the main purposes of the program is to "package" these activities as part of a broader, coordinated

planning and action effort, so that they might be mutually supporting and more integrally related to local community conditions and needs. Although certain maintenance elements of the program might be the focus of specific evaluation efforts, the program is directed to several opportunity goals, and a design for program evaluation would need to reflect corresponding direction, breadth, and openess.

Other examples of Type III are community action programs, economic planning and development programs, and some of the newer education programs directed at special problem groups or areas. In Alaska, the OEO-funded Community Enterprise Development Corporation program (CEDC) clearly falls within this category. It provides seed capital and some operating subsidies, technical assistance, and training to some two dozen different local businesses in rural Alaska villages. These businesses may be engaged in production, marketing, service, consumer, and related activities. They include fish production and marketing co-ops, arts and crafts co-ops, village stores, and other forms of local enterprise. While the provision of jobs and income is an immediate objective, the program is intended to develop the skills, attitudes, and knowledge that individual Alaska Eskimos and Indians would need to enter the mainstream cash economy, if they choose, and it seeks to establish a basis for further community and regional economic development through the attraction of additional public and private investment from other sources.

It is apparent that there are several program objectives, that "outcomes" may be many and varied, and that results could be analyzed at several levels — the individual participant, the co-op membership, the business, the community, and the region. In sum, CEDC is intended to be a combined social and economic development program. It has "maintenance" elements within it (jobs and income), and its general thrust is toward greater economic and social "opportunity" on the part of individuals, groups, and whole communities.

Type IV: Broad Aim — Maintenance Programs

Broad aim—opportunity programs often have concrete maintenance elements; some to the extent that it is useful for analysis to classify the maintenance aspects as a fourth type. The Community Enterprise Development Corporation (CEDC), for example, offers jobs and potentially higher incomes or increased savings to members of local cooperative organizations. Similarly, the Community Action Program (CAP) provides jobs and income directly within local CAP organizations and through the employment programs they sponsor in their communities. And the Urban Renewal Program of the Department of Housing and Urban Development (HUD), although justified as a means of revitalizing city centers, includes immediate and tangible benefits to businessmen (through subsidization of land purchases and clearing) and to the occupants (mostly middle-income and up) of new apartment houses constructed on cleared land.

A more explicit maintenance activity would be HUD's low-income public housing program, although this, too, especially in earlier years, was justified as an opportunity program, a means for changing the lives and behavior of low income families. While the public housing program might be evaluated in maintenance terms (e.g., whom does it reach and at what cost?), a similar maintenance-oriented assessment of urban renewal probably would not be welcomed by HUD officials since this could undermine the program's stated rationale. Several independent studies (not "program evaluations" required or requested by the federal agency) of urban renewal have in fact done this, to the discomfiture of HUD and its clientele, for they have shown that program benefits have tended to flow to the well-to-do, while the poor tend to bear the costs (e.g., demolished low rent housing, forced relocations).[19]

[19]See, among others, Herbert J. Gans, *People and Plans* (New York: Basic Books, 1968); Martin Anderson, *The Federal Bulldozer* (Cambridge: MIT Press, 1964); Scott Greer, *Urban Renewal and American Cities* (Indianapolis: Bobbs Merrill, 1965); and Charles Abrams, *The City is the Frontier* (New York: Harper and Row, 1965).

Thus, evaluation of opportunity programs may involve studies of maintenance elements, but these may be considered quite secondary, beside the point, or even harmful by program officials who feel compelled to justify and defend their programs in opportunity terms.

ALTERNATIVE APPROACHES TO EVALUATION

As has been stated, the first prerequisite to assessing the need for evaluation research and the form it should take is clear understanding of the nature and purposes of the program to be evaluated. Not all programs necessarily need special evaluation beyond what responsible officials themselves are able to do on the basis of adequate data collection and reporting routines built into the administrative system. Yet, many programs at one time or another may well benefit from an outsider's presumably more objective view, which may be supported by special research skills, social science or other professional knowledge, and program administration experience beyond the scope of specialist administrators.

If the purpose of the program is specific and concrete, the program will be easier to evaluate, and the evaluation research design may be tighter, more systematic, and possibly more "experimental." On the other hand, when program aims are broader and more elusive, then evaluation work becomes increasingly difficult in the sense that a tightly controlled, experimental study design is in most cases unlikely to be workable or appropriate.

But, even where objectives are clear and results simple to measure, a study design focusing on program effects may not be what is needed and wanted. For example, in a specific aim—maintenance program such as the food stamp program, the main question may not be "what is the program achieving?" Instead, it might be more appropriate to investigate whether the initial problem still exists, whether it continues to exist in the same form, whether the population has changed, and to determine who is benefiting from the program and who is being left out, at what cost in relation to benefits is the program operating, and what additional (unintended) effects is the program having that may help or hinder realization of other desirable ends. These are program

evaluation questions only in a very broad sense. They focus not on "outputs" or the measurement of results, but on wider program purposes and experience. Evaluation research designs may thus incorporate a variety of different approaches and skills, depending on what policy makers most need or want to know about the program.

Thus, some of the broader purposes that "evaluation" might serve include developing information about the changing extent and incidence of the problem, problem causes, characteristics of the population being served, side-effects of the program, and program efficiency. Additional "evaluation" work, broadly conceived, might include studies of public knowledge and attitudes concerning the program and its impact, of administrative processes and the relationships between administrators and clientele groups, and of the assumptions and logic underlying program objectives. Any or all of these factors (and there are potentially many others) could be critical to program performance and, hence, effectiveness in a given case.

Yet, granting all of the above, what if it is evaluation as the measurement of program results that is still wanted and needed? In such cases, it is essential that administrators and evaluators come to clear common understandings about what is expected to happen and when as a result of the program before starting an evaluation project. If this is not done, the chances are that they will become increasingly frustrated with each other, and both will be dissatisfied with evaluation results. It should be evident that the risks of this happening increase as program aims become less specific and as the time-lag between the application of program treatment and the accomplishment of the ultimate objective increases. This is because broader aims and uncertain relationships between program means and ends inevitably introduce ambiguity into the conception of what a program intends to do, and how what is done leads to the results sought. This is most likely to be the case with broad aim—opportunity programs.

There are two major alternative approaches to the problem of evaluation design in the case of complex opportunity programs. One emphasizes precision and sharp focus on selected program purposes and impacts. Rather than attempting to cover the full (and unspecifiable) range of

social and economic effects that might be associated with, for example, a community enterprise development project, evaluation might concentrate on decision-making patterns within the structure of the local organization in order to determine the relative roles and influence of funding agency officials, co-op board of directors, manager, and membership. The general purpose would be to develop indicators and measurements of organizational autonomy, given the broad social development objective of local self-sufficiency. Such a focus might then be complemented by parallel studies of financial status and managerial performance, viewing the organization strictly as a business enterprise. There would be no necessary relationship between the social-political and business-economic studies; it would be the task of program administrators to integrate such findings into their own broader comprehension of the program as a whole. Obviously, what may be gained in precision for evaluation research will be lost in comprehensiveness, insofar as total program experience and impacts are concerned.

The second approach would be "concerned with what form the action-program actually took, and with the details of its interaction with its surroundings, from which may be formed an inductive assessment of consequences."[20] Evaluators would be given maximum freedom to assess program developments and impacts, deciding in the course of their work what does and what does not deserve emphasis or concentration. It has been argued elsewhere that this may, in fact, be the only potentially effective way of assessing overall program performance and effectiveness in the absence of specific program objectives clearly traceable in program operations to expected and intended results precisely stated in advance — conditions virtually impossible to meet in broad aim - opportunity programs. Evaluation in such cases "would be much more concerned with learning than with measuring."[21]

The two approaches are not mutually exclusive; indeed, they can be mutually reinforcing. But even relatively large-scale programs can be "over-evaluated," with teams of

[20] Weiss and Rein, *op. cit.*, p. 142.

[21] *Ibid.*

social researchers, specialist-technicians, and others trying to do too much, too soon, and in the process disrupting program administration and alienating program administrators. The approach taken will depend largely on what the program agency wants from evaluation research, what it is willing to pay for, and how much confidence it is willing to place in evaluation researchers. If the agency wants greater control and accountability, then the research design must be narrower and more specific, but the risks are thereby increased that research findings will be incomplete and fragmented and may even miss the point of the program. If the agency is willing to "trust" evaluators, allowing much more improvisation in an open-ended research design, then it will be more likely that evaluation results will not be tailored to their specific needs for information or directly relevant to program adminstration, whatever may be gained in deeper understanding of program policy, experience, and impacts.[22]

Given the present state of the art, and the demands placed on evaluators for "answers," the available alternatives may not be reassuring to agencies required to "prove" the effectiveness of their programs through evaluation research.

As has been repeatedly emphasized, the single most important rule of evaluation research method, whatever the program, is that policy makers and administrators clearly define what it is they want to know. Prospective evaluators should work closely with them in formulating the problem. This is essential not only so that evaluators can respond intelligently to requests for program research, but, as important, so that there may be some basis for determining whether an evaluator or evaluation team possesses the skills and resources necessary to deal with, say, community survey work, management analysis ("efficiency studies"), or broad assessment of the incidence and extent of the problem to which the program is directed.

[22] A relevant and very sensible suggestion is that "if the policy-maker simply desires advice as to what he should do, he had better rely on the intuition of a man of wide experience and demonstrated understanding rather than on the intellectual skills and techniques of the social scientists." (Max F. Millikan, "Inquiry and Policy: The Relation of Knowledge to Action," *The Human Meaning of the Social Sciences* [New York: Meridian Books, 1959], p. 165.)

CONCLUSION

Evaluation research conducted under current conditions risks one or both of two types of failure. The first is a technical failure of evaluation research design that results from the attempt to define specific program objectives and measure specific program results that do not correspond to the actual breadth and scope of program aims. Such research provides incomplete and partial accounts of program effects. Consequently, administrators and other defenders of the program may charge that the evaluation research has underestimated or "missed the point" of the program. The second type of failure is political in character -- the research may indicate that the program is very weak and ineffective, and thus produces findings that are unacceptable to policy makers and administrators, who may have oversold the program in the first place and who have a major stake in program success, or at least the appearance of success.

Program evaluation research will continue to be a precarious and sometimes futile undertaking so long as (1) evaluators attempt to focus on program effects that are either minimal or peripheral to broad (or even vague) objectives, as conceived by administrators, and (2) administrators and policy makers have a significant political stake in any assessments of program effectiveness. Evaluation research can best search out and measure effects when programs have specific aims and are reasonably full-scale attacks on reasonably well-defined problems, and it can have an impact on decision-making when administrators insist less on "acceptable" findings and more on objective understanding of the determinants of program effectiveness. Neither condition appears to prevail today.

This does not mean that evaluation research should not be done. Rather it indicates that the research should be designed to match more closely the scope and nature of the programs to be evaluated, and that it should be done with fuller awareness on the part of both administrators and researchers of the political and technical problems likely to be encountered. In some cases, program effects may be a secondary consideration, and emphasis may be placed on the evaluation of objectives, means, and their "fit" to widely

varying social and economic contexts and problem situations. In other cases, fairly well defined programs can be evaluated in more conventional ways, with a more direct focus on effects. But, by and large, evaluation research will often need to take "softer" forms than are preferred by the more exacting methodologists among social researchers. It will also need to include policy and problem analysis within its scope, recognizing that what administrators can accomplish is in large part determined by the character and quality of the policy making processes that led to the establishment of their programs in the first place.

REFERENCES

Campbell, Donald T., "Considering the Case Against Experimental Evaluations of Social Innovations," *Administrative Science Quarterly*, Vol. 15 (March 1970), pp. 110 - 113.

Campbell, Donald T. and Stanely, Julian C., *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally and Company, 1963.

Caro, Francis G., "Approaches to Evaluative Research: A Review," *Human Organization*, Vol. 28, No. 2, (Summer 1969), pp. 87 - 99.

Carter, Reginald K., "Clients' Resistance to Negative Findings and the Latent Conservative Function of Evaluation Studies," *The American Sociologist*, Vol. 6 (May 1971), pp. 118 - 124.

Cicirelli, Victor G., and others, "The Impact of Head Start: A Reply to the Report Analysis," *Harvard Educational Review*, Vol. 40, No. 1, (February 1970), pp. 105 - 129.

Cohen, David K., "Politics and Research: The Evaluation of Large-Scale Social Action Programs in Education," manuscript of article appearing in March 1970 issue of the *Review of Educational Research*.

Ferman, Louis A., "Some Perspectives on Evaluating Social Welfare Programs," *The Annals*, Vol. 385, (September 1969), pp. 143 - 156.

Freeman, Howard and Sherwood, Clarence, "Research in Large-Scale Intervention Programs" *Journal of Social Issues*, Vol. 21, No. 1, (January 1965), pp. 11 - 28.

Hayes, Samuel P., Jr., *Evaluating Development Projects*, 2nd ed. Paris; UNESCO, 1966.

Hyman, Herbert H. and Wright, Charles R., "Evaluating Social Action Programs," *The Uses of Sociology*, Paul F. Lazarsfield, et al., eds., New York: Basic Books, 1967.

Hyman, Herbert H., Wright, C., and Hopkins, T., *Application of Methods of Evaluation*, Berkely: University of California Press, 1962.

Mann, John, *Changing Human Behavior*, New York: Scribners, 1965.

Marris, Peter and Rein, Martin, *Dilemmas of Social Reform*, New York: Atherton Press, 1969, Chapter 8: "Research."

Miller, S. M., "Evaluating Action Programs," *Transaction*, Vol. 4 (June 1967).

Millikan, Max F., "Inquiry and Policy: The Relation of Knowledge to Action," *The Human Meaning of the Social Sciences*, New York: Meridian Books, 1959.

Madge, John, *The Tools of Social Science*, New York: Anchor Books, 1965, Chapter 5: "Experiment."

Moynihan, Daniel P., *Maximum Feasible Misunderstanding*, New York: The Free Press, 1969, Chapter 8: "Social Science and Social Policy."

Myrdal, Gunnar, *Objectivity in Social Research*, New York: Pantheon Books, 1969.

National Science Foundation, *The Nation's Use of the Social Sciences*, Report of the Special Commission on the Social Sciences of the National Science Board, Washington: Government Printing Office, 1969.

Rein, Martin, "Social Policy Analysis as the Interpretation of Beliefs," *Journal of the American Institute of Planners*, Vol. 37 (September 1971), pp. 297 - 310.

Rodman, Hyman and Kolodny, Ralph, "Organizational Strains in the Researcher — Practitioner Relationship," *Human Organization*, Vol. 23, (Fall 1964), pp. 171 - 182.

Rossi, Peter, "Evaluating Social Action Programs," *Transaction*, Vol. 4 (June 1967).

Schick, Allen, "From Analysis to Evaluation," *The Annals*, Vol. 394 (March 1971), pp. 57 - 71.

Smith, Marshall S. and Bissell, Joan S., "Report Analysis: The Impact of Head Start," *Harvard Educational Review*, Vol. 40, No. 1 (February 1970), pp. 51 - 104.

Suchman, Edward A., *Evaluative Research*, New York: Russell Sage Foundation, 1967.

Weiss, Robert S. and Rein, Martin, "The Evaluation of Broad-Aim Programs: A Cautionary Case and a Moral," *The Annals*, Vol. 385, (September 1969), pp. 133 - 42.

Weiss, Robert S. and Rein, Martin, "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties, and an Alternative," *Administrative Science Quarterly*, Vol. 15 (March 1970), pp. 97 - 109.

Wholey, Joseph S., and others, *Federal Evaluation Policy: Analyzing the Effects of Public Programs*, Washington, D.C.: The Urban Institute, 1970.

Williams, Walter and Evans, John W., "The Politics of Evaluation: The Case of Head Start," *The Annals*, Vol. 385 (September 1969), pp. 118 - 32.

END