

Coastal Zone
Information
Center

10625

**COASTAL ZONE
INFORMATION CENTER**

AUG 19 1977

PROGRAM MANAGEMENT
AND
THE FEDERAL EVALUATOR

Pamela Horst
Joe N. Nay
John W. Scanlon
Joseph S. Wholey

162-0010-6



An Urban Institute Reprint

Urban Institute
H
62
.H67
1974

The Urban Institute is a nonprofit research corporation established in 1968 to study problems of the nation's urban communities. Independent and non-partisan, the Institute responds to current needs for disinterested analyses and basic information and attempts to facilitate the application of this knowledge. As part of this effort, it cooperates with federal agencies, states, cities, associations of public officials, the academic community and other sectors of the general public.

The Institute's research findings and a broad range of interpretive viewpoints are published as an educational service. Conclusions expressed in Institute publications are those of the authors and do not necessarily reflect the views of other staff members, officers or trustees of the Institute, or of organizations which provide funds toward support of Institute studies.

These research findings are made available in three series of publications: Books and Reports, Papers and Reprints. A current publications list is available on request.

Publications Office

The Urban Institute
2100 M Street, N.W.
Washington, D.C. 20037

REFER TO URI-10115 WHEN ORDERING.

162-0010-6 "PROGRAM MANAGEMENT AND THE FEDERAL EVALUATOR," by Pamela Horst, Joe N. Nay, John W. Scanlon, and Joseph S. Wholey, is reprinted with permission from the *AMERICAN SOCIETY FOR PUBLIC ADMINISTRATION*, appearing in *PUBLIC ADMINISTRATION REVIEW*, July/August, 1974.

B/77/300

COASTAL ZONE
INFORMATION CENTER

PROGRAM MANAGEMENT
AND
THE FEDERAL EVALUATOR

Pamela Horst
Joe N. Nay
John W. Scanlon
Joseph S. Wholey

162-0010-6

U. S. DEPARTMENT OF COMMERCE NOAA
COASTAL SERVICES CENTER
2234 SOUTH HOBSON AVENUE
CHARLESTON, SC 29405-2413

August 1974



Property of CSC Library

THE URBAN INSTITUTE
WASHINGTON, D.C.

Urban Institute
H62 .H67 1974
2391298

FEB 19 1997

P ROGRAM MANAGEMENT AND THE FEDERAL EVALUATOR

Pamela Horst, Joe N. Nay, John W. Scanlon, and Joseph S. Wholey, *The Urban Institute*

In 1969, The Urban Institute completed an extensive study of federal evaluation and concluded that, "The most impressive finding about the evaluation of social programs in the federal government is that substantial work in this field is almost non-existent."¹ A limited resurvey of the field in 1972 revealed a quite different picture: funds committed to evaluation had mushroomed, many studies had been completed, and the use of large-scale social experimentation was increasing.²

This growth in evaluation has contributed information—often imperfect, sometimes incorrect—to today's arguments about the direction, method, and purpose of social programs. Without evaluation, many arguments would have remained at the level of polemic. There is no question that the presence of program evaluation has heightened the consciousness of federal program managers and policy makers to the fact that they may, from time to time, have to respond to queries about the effectiveness of their programs.

While evaluation has firmly established itself since 1969 in both the budget and the administrative rhetoric of the federal government, there is little evidence to show that evaluation generally leads to more effective social policies or programs. On the contrary, the experience to date strongly suggests that social programs have not been as effective as expected and have not improved in

performance following evaluation. This situation can be phrased as a critical management problem which we see confronting government agencies:

Why have those in charge of programs and those who evaluate them not been able to join their efforts in a way that leads more frequently to significant improvements in program performance?

Having been able both to observe and to participate in the development of federal program evaluation, we have chosen here to raise three propositions about the root causes of the above problem. If these causes are the crucial ones and if we can come to understand their true impact, federal program management and evaluation stand on the edge of a period of increasing success. If not, and the causes of these weaknesses continue to be ignored, then evaluation, program management, the programs themselves, and those the programs are intended to serve will all continue to suffer.

This paper elaborates on why the three root causes, when they exist, block further improvement of many programs. The idea of a "preassessment" of program evaluability is introduced as one tool for improving both program management and program evaluation. We begin with a discussion of the conventional treatment of evaluation problems

and then present an alternative diagnosis and prescription. Although much of the material presented is addressed to federal managers and federal evaluators, we believe the problems and solutions discussed also hold for state and local government.

Apparent Causes of Evaluation Problems— and an Alternative Statement

Most reviews made to determine what causes programs and their evaluations to be ineffective include one or more of the following conclusions:³

- Evaluations are not planned to support decision making.
- The timing, format, and precision of evaluation studies are not geared to user needs.
- Evaluation findings are not adequately communicated to decision makers.
- Different evaluations of the same program are not comparable.
- Evaluation fails to provide an accumulating, increasingly accurate body of evidence.
- Evaluation studies often address unanswerable questions and produce inconclusive results.

The first three apparent causes deal with aspects of evaluation use. They occur at the interface between the producers of evaluations and the prospective users of evaluation.

The second three apparent causes deal with the methods used by the evaluators in assessing the interventions of the programs in society. They occur at the interface between the producers of evaluation and the program as it exists. They concern flaws in making measurements and comparisons and in drawing conclusions.

Our experience to date in studying the management problem—namely, the lack of significant improvement in program performance—and in watching various agencies attack the apparent causes of the problem has led us to conclude that these six statements largely refer to symptoms, rather than causes. We believe that the causes of the problem may more properly be described by one or more of the following three propositions concerning the program itself:

- *Lack of Definition:* the problem addressed, the program intervention being made, the expected direct outcome of that intervention, or the expected impact on the overall society or on the problem addressed are not sufficiently well defined to be measurable.

- *Lack of Clear Logic:* the logic of assumptions linking expenditure of resources, the implementation of a program intervention, the immediate outcome to be caused by that intervention, and the resulting impact are not specified or understood clearly enough to permit testing them.
- *Lack of Management:* those in charge of the program lack the motivation, understanding, ability, or authority to act on evaluation measurements and comparisons of *actual* intervention activity, *actual* outcomes, and *actual* impact.

When one or more of these three propositions is true, both the problem (lack of significant improvement in program performance) and the six apparent causes listed earlier can easily occur. In cases where the first two propositions hold, an enormous range of possibilities will present themselves as to which measurements and comparisons to make—with no criteria for making sound choices. In cases where the last proposition holds, even exceptionally high quality evaluation is not likely to be used well, if used at all. If a program suffers from one or more of these three flaws, there is a very low probability that evaluation information useful to program improvement can be produced. Thus the program may be “unevaluable” until the flaws are corrected.

A statement that the quality and value of evaluation are strongly affected by the degree to which these three conditions exist is not a startling finding. What has not been realized or acknowledged in the past, however, is that these three factors are not the responsibility of the evaluator. While the conventional apparent causes relate to how the evaluator does his job, these latter three propositions describe an organizational environment over which the evaluator typically has little control. Evaluators, more than any other group in an agency, will appear unable to complete their work successfully when these conditions exist, regardless of how they deal with the apparent causes.

Why the Apparent Causes Are Suspect

In the past few years, we have conducted a number of federal program evaluations and helped to develop evaluation planning systems for several federal agencies. In the course of our work, we have observed many attempts to treat the six apparent causes directly by improving the use and

methodology of evaluation. These attempts include policy review and dissemination panels, letting contracts for methodology development, high level reviews of evaluation plans, task forces to select better questions for evaluation, better systems for collecting data, requiring program offices to submit advance descriptions of how evaluation findings will be used, and the tightening of contract selection and monitoring procedures to increase contractor responsiveness to agency needs. In some cases, the "solution" was reorganization: centralizing previously decentralized evaluation units. Since revenue sharing, talk of decentralizing a previously centralized evaluation office has gained popularity. In this case, the headquarters office would no longer be responsible for conducting national program evaluation, but instead would go into the business of building local evaluation capability.

As these proposed solutions were implemented, however, we have continued to talk with and work with participants in the process from the assistant secretary level, through the program level, and on down to the recipients of services. We find that the management problem—that is, the lack of significant improvement of program performance—continues to exist and the same apparent causes continue to be cited, whether there are high or low quality evaluation efforts. Improvements in programs and in delivery of effective services remain far below the levels desired or expected. If the root causes of the problem lay within the evaluation process, we believe that these correctives would be showing some degree of success. Consequently this experience led us to search for alternative explanations and, finally, to the consideration of the three conditions stated above as root causes of the problem.

The Source of the Problem

The significance of the proposed causes can best be understood by contrasting the nature of the intervention that the social programs of today attempt to make in the society at large with the principal types of program interventions attempted in the past. Many older, classical government activities involved program interventions whose nature was clearly defined and agreed upon and which were described in detail in a body of law or regulation (e.g., Social Security). The implementation of such activities was largely an

act of administration of the laws and regulations. Evaluation of success or failure of the act of implementation was primarily a matter of assessing compliance with the guiding laws and regulations. Discretion was at a minimum (at least over the short term). Arguments might take place about whether *goals* were adequate, but the *details of the program intervention* were determined in advance.

In contrast, many new missions that the federal government has been called upon to undertake (e.g., lowering hard core unemployment) involve problems in which the proper program intervention mechanism is not well understood, or defined, or in some cases even known. Since in these cases no one knows exactly what detailed program intervention will be of value, greater management discretion is allowed and exercised. While in some cases research may be undertaken or experiments may be made to increase understanding, more typically a purportedly successful type of program intervention is simply put into place and an agency or bureau is charged with making it into a successful operation. In this case, evaluation is expected to report to those in charge of a program on whether the use of discretion in choosing specific program intervention techniques was successful and perhaps to suggest modifications or alternatives.

The newer program areas are characterized by uncertainty and discretion: uncertainty as to the nature of the problem and what constitutes effective strategies of intervention, and discretion in how the problem and the intervention are defined and how the intervention is implemented. These conditions make sound and rapid evaluation all the more important to effective management. Consider, however, how today's program environments can disable evaluation through three factors: lack of definition, lack of a clear logic, and lack of management.

Lack of Definition

Examination of program legislation, regulations, policy manuals, plans, and budget to determine what a program intervention is can be very deceptive. What at first seems clear often evaporates when the test of measurability is applied. The language used turns out to be ambiguous precisely where it would have to be specific in order for evaluation to be useful. Three common forms of inadequate language are: the vaporous wish, local project packaging, and how-to-do-it rule making.

The vaporous wish is the eloquent but elusive language of goals put forward for most federal programs. Exactly what are the "unemployability," "alienation," "dependency," and "community tensions" some programs desire to reduce? How would one know when a program crossed the line, successfully converting "poor quality of life" into "adequate quality of life"? Would anyone recognize "improved mental health," "improved local capability," or "revitalized institutions"? The problems addressed by social programs are almost never stated so that institutions, people, or the relevant socioeconomic conditions could be classified according to the degree to which they are afflicted with a problem. It is very hard to propose a solution to a problem that is ill-defined or undefined. How much harder it is to evaluate the success of that proposed solution.

Next, there is the project packaging language which purports to describe the intervention activity to be planted in the field and the expected outcome for those directly served by that activity. As any experienced site visitor will attest, this language is often so annoyingly imprecise that it is difficult to tell what parts of a local operation are under discussion and even harder to distinguish compliance or assess performance. For example, project characteristics prescribed in various program guidelines include: "coordinative mechanism," "integrated services," a "range of modalities," "extended career ladders," "accessibility of services," "continuity of care," "multi-disciplinary teams," "outreach capability," etc. Projects should produce "upgraded job skills," "increased cultural enrichment," "increased personal autonomy," "improved family cohesion," etc. Rarely are useful measures or norms for these activities and outcomes provided.

How-to-do-it rule making is the third kind of language that is commonly found. Here the terms are very concrete and specific. We find guidance on factors like the qualifications of project directors, the contents of affiliation agreements with other local agencies, reporting relationships, the use of consultants, and accounting practices. This guidance appears to be definite and all inclusive. Closer examination shows that it usually tells how to run the part of a project which does not deal directly with the intervention into society. Guidance for the part of the project which actually produces effects in society is not provided.

When these three forms of language predominate, the intervention activities in the field may be diverse indeed. Our experiences examining field operations indicate that program packaging is generally skin deep and that very different project activities and definitions of outcome often parade under the same assumed program names. An examination of 20 projects in the same program will often reveal 20 very different program intervention designs, different in activity and purpose. This means that the program activity and objective, as implemented in the field, cannot be defined on a common base of measurable terms. It is often difficult to find any consensus among federal level policy makers as to what the definitional base should be. This lack of a common framework can disable management and evaluation efforts alike.

It is becoming clearer that many federal social programs are simply envelopes for a large federal investment in a problem area. A program may be deceptive in the sense that it has enough content to allow it to be described in the media, lobbied into existence, and established as a federal effort — and yet the program interventions are not spelled out in any detail. Many program administrators over the last decade have essentially received a program envelope with only vaporous wishes and money inside. Although more detailed definition may not have been necessary in order to spend the money, much more detailed definition is needed to evaluate the process and outcome.

If it is decided that certain programs should be further defined, who in an agency should be responsible for the tasks? It should not be left to the auditors or to the evaluators or to the information system people, because the choice of specific measurable definitions is not merely a technical task. The definition of what is to be measured in a program is central to policy making and program management. If there are many different ways to measure the problem a social program purports to influence, this often means that there are many different problems. For many programs, no one has yet exercised the prerogative of selecting which specific set of social ills the program is trying to cure or the methods of cure. Legislation or regulations rarely make this choice, and the choice has policy implications since it further specifies program intent and intervention. One of the major factors in shaping and directing a program is carefully selecting what the program is

going to do. The failure to define measurable interventions, outcomes, and impact for a program is a major policy making defect. Those in charge of the agency and the program, rather than the evaluators, should have primary responsibility for program definitions.

Lack of a Clear Logic of Testable Assumptions

Even if the policy makers or program managers have provided measurable definitions, there still may not be unanimity within a federal social program about design or logic. As a result, different evaluation efforts are often based on *different* assumptions linking program intervention with immediate outcome and ultimate program impact. The measures and data collection instruments used are those that seem most reasonable to the evaluator. In this context it is easy to understand why evaluation findings are often noncomparable. When there is no carefully determined framework to guide the program, there is, of course, no such framework for evaluation studies. Nor is there a framework for systematically accumulating knowledge of program performance. In fact, it becomes unclear what program performance means.

Program assumptions might be as simple as that "the transfer of money to school districts will raise the reading level of disadvantaged students" or that "the training of the unemployed will lower unemployment." Often the board program charters from the Congress referred to earlier have caused clusters of competing assumptions to grow up in many social programs. One set of assumptions may be used for arguments with friends, for instance, and another for argument with enemies. This may be good politics, but it makes for difficult evaluation design, since evaluation design should relate to the information needed to validate, refute, or modify a set of operating assumptions.

Without an adequate description of the assumptions governing the intervention of a program into society, it is more likely that evaluators will be asked to address unanswerable questions far removed from the actual activities taking place. To take a quite reasonable example, a program office might insist on funding an evaluation to assess the relative effectiveness of different drug treatment modalities. The evaluator may then find that these modalities do not represent pure, mutually exclusive approaches which are replicated in multiple

local settings. He is likely to find, on the contrary, that a "halfway house" or a "therapeutic community" in one locale bears no resemblance in operating assumptions to others which go by the same name. After spending a lot of money, time, and effort, the evaluator will be forced to tell the agency what types of programs are really out there, rather than how successful they are, and also that the only way to test the effectiveness of alternative assumptions of treatment is to implement a program-level experiment, or introduce planned and enforced variations into the program design. Those in charge of the program may feel that the evaluator has once again failed to answer their questions. There are many examples of evaluations being mounted to answer questions which bear no relationship to the program activity actually taking place in the field. This counterproductive practice results from the failure of the agency to describe carefully the program assumptions so that they can be implemented and tested.

Summing up, even when the intervention, expected outcome, and impact are defined in measurable terms, the more subtle questions of the logic linking (a) program expenditures to production of the intervention, (b) intervention to outcome, and (c) outcome to impact on the problem must still be considered. The use of the word "logic" here is not meant to imply that the linking assumptions are loose or tight, valid or invalid, defensible or stupid. All that is implied is that a program in reality is based on an interrelated set of assumptions about what is believed to happen (and sometimes why) when money is spent and the intervention made. The absence of statements of these assumptions might be expected to cause a problem for both program managers and evaluators. The evaluators often notice the absence first, however, because they must design tests of these assumptions. Tests cannot be designed for people who are unable to, or refuse to, state their assumptions.

Once again (as with measurable definitions) the statement of the logic of testable assumptions is a policy question, not one that should be decided by the evaluators. Evaluators should test the assumptions about what works. Those in charge should make the initial assumptions underlying the funding and operation of the program.

Lack of Management

To get at the significance of lack of manage-

ment, it is important to realize that evaluation is useful only if it is, in fact, a tool of management. A manager has a variety of tools to employ that include direction of his line management, planning, budgeting, audit and financial control, administration (for that part of his activity that can be clearly defined and where a law or set of rules is used to guide program implementation), policy analysis, and evaluation. Evaluation is needed principally in support of policy analysis and management discretion. Evaluation performs the same function for management that audit and control do for budgeting and that compliance checks do for administration.

One way of understanding the role of evaluation as a management tool is to explore how a "textbook manager" might use evaluation in attempts to improve program performance and then to contrast that with the way evaluation more frequently is used.

Evaluators and the "textbook manager" cooperate very well. When the policy decisions about program design are to be made, the evaluator asks the manager to specify the measurable definitions, the assumptions of the program linking these definitions, what kind of performance data would cause the manager to act, and the kinds of action the manager has the authority and willingness to implement. Armed with this guidance, the evaluator estimates the level of error associated with collecting the evidence, estimates the ranges of possible findings, and bounds the cost of the proposed evaluation. The evaluator is then equipped to provide a service not commonly rendered at present. He can advise management on the cost and feasibility of procuring evaluation evidence, and the manager can weigh these factors against the potential value of evidence for improving program performance. When the evaluation is finally commissioned, the evaluator has a clear basis for judging the best level of aggregation, precision, and delivery schedule because he has a user for the proposed evaluation. Many market surveys and internal evaluations are conducted this way in industry. When this kind of rational planning occurs, one does not generate evaluation studies in search of users and uses.

The utility of social program evaluation depends at least in part upon defining the decision context as well as the program design. The "textbook manager" has already defined his program in measurable terms and has indicated what

it purports to accomplish. If evaluation is to contribute to program improvement, there must be at least a few decision areas where the manager will rely on program performance feedback (measures of impact, outcome, intervention activities), as well as on political pressures, popular approaches, or his own hunches and beliefs. Else why buy evaluation at all? The "textbook manager" knows in advance and can specify what level of evidence will prompt him to act at all, or cause him to select among alternative actions. Further, he has the authority to act.

Return with us now to reality, where the typical government administrators live. These administrators participate in continual agency debate over program issues, but the debates proceed in a language which means different things to different people. The debates are not centered on a measurable set of program descriptions nor are the assumptions guiding the program intervention made clear enough to be testable. In fact, most of the people in this world will go to great lengths to keep these two things ambiguous in order to expand their area for maneuver. The administrator is a decision maker—he does take action. As in "textbook" management, many of his actions are based on guesses about what is needed, shifting academic opinions and political support, and the demands of a set of higher level policy makers subject to continual turnover. Unlike the "textbook" management, however, the typical government administrator does not establish and test assumptions linking intervention activities to program performance. Typical government administration might be called "pseudo-management," because all its management activity takes place in a process that is not linked to actual program results. In its own terms, such "pseudo-management" is good if its activities remain acceptable to an ever-changing cast of characters at the policy level.

Evaluators and pseudo-managers operate independently of one another. There is no basis for communication between them. The pseudo-manager has no real use for evaluation and the evaluator can provide few, if any, services to assist in pseudo-management. In fact, sound evaluation results may present a clear and present danger to the pseudo-manager. In this environment, the evaluator can expect his work to have minimal impact. The problem for the evaluator is to distinguish pseudo-management from textbook

management. On the surface it appears to us that pseudo-management predominates in social agencies; the potential for textbook management is yet unknown.

Our emphasis on identifying actual users of evaluation and on pre-specifying the decision context and uses of evaluation information may seem excessive. Yet the desired use of evaluation information determines not only how much it is worth but also the form and accuracy that it must have. And if those in charge of a program have no use for information about that program, then there is no real way to design an adequate evaluation for them. What this might mean may be demonstrated by an example.

Assume that a federal drug treatment program for heroin abusers defines outcome success in the following terms: the client reveals absence of heroin use six months after discharge from treatment, as tested by three randomly spaced urinalyses during the follow-up period and one urinalysis at the end of the six months. Those in charge of the program say that they require information about this outcome to assist in decisions about the following: allocation of technical assistance among drug treatment projects, reallocation of funds among projects, and assignment of headquarters staff to study problems associated with achieving a desirable outcome level. But suppose those in charge are challenged to specify in advance how decisions might vary with the range of possible evaluation findings. For example, will a task force convene for program redesign if national program cure rates average 5 per cent, 15 per cent, or 50 per cent? Will technical assistance be given to projects whose average cure rate falls below 5 per cent? Is there technical assistance to give? Can projects be closed down? Will a stated national objective of a 30 per cent cure rate be adjusted downward, if the actual average cure rate found is 15 per cent? This type of dialogue would permit the evaluator to assess the potential value of evaluation information by identifying plausible and practical uses for it and also permit the evaluator to assess the specific type and accuracy of the information required.

The level of validity and reliability required in measurable data should be an important factor used in analyzing the method of collection, the cost of data collection, and the methods and cost of data analysis before data collection efforts ever begin. The "conclusiveness" of data only takes on

meaning in relation to particular actions the data may suggest. But as we saw earlier, if everything is left ambiguous, no one will know what level of evaluation findings would or should prompt action and therefore what level of validity and reliability are required in the evaluation data. This means that, in our example, drug program evaluations which show cure rates of 2 per cent, 5 per cent, 20 per cent, 50 per cent, and 75 per cent could all be dismissed by the pseudo-managers as "inconclusive" for decision making.

When a single individual does not have the authority to take or to elaborate on the kinds of action mentioned above, those individuals whose consensus is required must be found and consulted. The point is that management of a program is a policy matter. Evaluation cannot prescribe management actions. Rather, the needs of management should define evaluations.

The Consequences of Evaluating When These Conditions Exist

Why should the evaluator worry about the soft, unmeasurable underbelly of social program goals, objectives, and activity; about the obscure logic of program assumptions; or about whether there is a management vacuum? If our analysis is correct, weaknesses in these areas can disable an evaluation effort while making the failure appear to be the evaluator's own doing.

If the agency evaluator, alone or with a contractor, attempts to carry out an evaluation of a program where these flaws exist, our experience indicates that there are two highly likely outcomes. First, the evaluator's attempt to define the program in measurable and logical terms will flounder. No available methodology can bridge the gap between the program as implemented in the field and the program as suggested by program goal statements. Thus the results of his evaluation are likely to be labeled "inconclusive," "abstract," or "an effort to develop methodology." Second, his findings will not be responsive to the information needs of those in charge of the program. He may produce the wrong information, or information that is too imprecise or too sparse. Even if the evaluation is technically unimpeachable, those in charge of the program may find it irrelevant to their decision context, seeing no way to act upon the information.

We have suggested that the definition of

measurable program design and of testable assumptions about how the program works is a major policy issue which should be resolved by policy makers and program managers within the discretionary boundaries of program legislation. Program policy making is not the job of the agency evaluator, and he should not undertake the task even if it is disguised as a "technical" choice of the proper program measures needed to conduct an evaluation.

Some Sources of Leverage

Is there a strategy that evaluators can adopt to return the jobs of policy making and program management to policy makers and evaluators—and improve the utility and yield of evaluation (and program) dollars? Fortunately, some factors in the present federal environment may supply the leverage needed to force attention to the three conditions (lack of definition, lack of clear logic, and lack of management) that have proven costly to program effectiveness and evaluation.

First, there is less naiveté about federal social programs today. More awareness exists that attacking a vague problem with an unproven social, behavioral, or economic theory is not likely to bring success. Raising issues about program definitions and assumptions is now more likely to strike a responsive chord in this climate. Secondly, the federal budget is not expanding rapidly, and the present Administration and the Congress are placing more emphasis on accountability. Third, both the Congress and citizens are pushing for more effective delivery of public services, and more evidence of effectiveness.

The evaluator, with some help from high level policy makers and program managers, may be able to take advantage of these potential sources of leverage and use them to force the definition that makes evaluation possible. At least he may assure that his efforts are expended in areas where there is the best chance of success. The tool that we recommend he employ is a "preassessment of evaluability" for every program that is a candidate for evaluation.

Preassessment of Evaluability

We recommend a process of pre-evaluation design.⁴ If conducted in proper detail, this process can provide what might be called a "rapid feedback evaluation" of the present status of a

program and its information base, and can make clear whether a major evaluation effort is or is not warranted. In essence, the three root causes of problems in program evaluation can be transformed into a set of criteria for determining the evaluability of a public program. These criteria are expressed in the following questions:

- Are the problems, intended program interventions, anticipated outcomes, and the expected impact sufficiently well defined as to be measurable?
- In the assumptions linking expenditure to implementation of intervention, intervention to the outcome anticipated, and immediate outcome to the expected impact on the problem, is the logic laid out clearly enough to be tested?
- Is there anyone clearly in charge of the program? Who? What are the constraints on his ability to act? What range of actions might he reasonably take or consider as a result of various possible evaluation findings about the measures and assumptions discussed above?

In a sense the criteria are sequential. Measurable definitions form a basis for the testable assumptions. Then both serve as a basis for the consideration of the range of decisions that those in charge of the program might make as a result of information about actual costs, interventions, outcomes, and impact.

In practice the evaluator will have to judge the degree to which the three criteria are satisfied for particular programs. The evaluator generally has several programs in his agency that can be evaluated at any one time. In initial planning, the evaluator should focus on testing each program against these three criteria, using the best information available from the programs themselves to assess how valuable each program may be. This assessment should be discussed directly with policy makers and program officials. The interaction between evaluator and program officials may assist policy makers and program officials to define the measures and specify the logic of assumptions that need to be tested.

The next task is to decide which programs meet all three criteria. Then programs that meet some criteria, or almost meet all criteria, may be sorted out. Finally, in most agencies, a third group of programs will emerge which satisfy few—if any—of the three criteria.

At this point the evaluator will have completed his own preassessment of the "evaluability" of the programs of his agency. It is almost useless to explore questions of use and methodology for programs that clearly do not meet the criteria. The next and final step is both a possible source of leverage for the evaluator and a somewhat risky business in many agencies.

Clearly Naming the Problem for Others

The evaluator has now created three lists of programs: "evaluable," "potentially evaluable with further program or management definition," and "not evaluable." Since these problems are now understood to involve policy and management questions, as well as evaluation design questions, the list has two uses.

First, the evaluator should evaluate only the programs that are evaluable. He should agree to help with the definitional problems of potentially evaluable programs. But he should not hesitate to name the nature of the problem. The evaluator should tell policy makers and program managers whether their programs are or are not evaluable, and why. Second, the evaluator should bring the

serious problems on the list to the attention of the top level of the agency hierarchy so they will know which programs are or are not evaluable, and why.

These actions may be very risky things to do in many agencies, but it can prevent a lot of useless evaluation attempts and later recrimination. We believe that they would force improvements in program performance as well.

Notes

1. Joseph S. Wholey, et al., *Federal Evaluation Policy*, (Washington, D.C.: The Urban Institute, 1970).
2. Garth N. Buchanan and Joseph S. Wholey, "Federal Level Evaluation," in *Evaluation*, Vol. 1, No. 1 (Fall 1972), pp. 17-22.
3. For a concise overview of the literature in which these criticisms have been put forward see Francis G. Caro (ed.), *Readings in Evaluation Research*, Russell Sage Foundation, 1971, pp. 9-15. In our own work we have had access to unpublished internal assessments of evaluation efforts by several federal agencies; the majority of these note agency dissatisfaction with their evaluation product and identify many of these apparent causes as major influences.
4. See John D. Waller and John W. Scanlon, *Urban Institute Plan for the Design of an Evaluation* (Washington, D.C.: The Urban Institute, March 1973).

OTHER SELECTED URBAN INSTITUTE PUBLICATIONS

BOOKS

Federal Evaluation Policy: Analyzing the Effects of Public Programs, Joseph S. Wholey, John W. Scanlon, Hugh G. Duffy, James S. Fukumoto, and Leona M. Vogt, 1970, URI 40001, Paperback, \$2.95

Urban Processes: As Viewed By The Social Sciences, Kenneth J. Arrow, James G. March, James S. Coleman, Anthony Downs, and William Gorham, 1970, URI 20001, Paperback, \$1.95

The Unemployment-Inflation Dilemma: A Manpower Solution, Charles C. Holt, C. Duncan MacRae, Stuart O. Schweitzer, and Ralph E. Smith, 1970, URI 60003, Paperback, \$2.95

Blacks and Whites: An Experiment in Racial Indicators, Michael J. Flax, 1971, URI 60002, Paperback, \$1.50

New Towns In-Town: Why a Federal Program Failed, Martha Derthick, 1972, URI 70006, Paperback, \$2.95

Public Prices For Public Products, Selma J. Mushkin, Editor, 1972, URI 90009, Paperback, \$6.50; URI 90010, Hard cover, \$10.95

What is Revenue Sharing?, Charles J. Goetz, 1972, URI 13000, Paperback, \$1.95

Forecasting Local Government Spending, Claudia DeVita Scott, 1972, URI 50010, Hard cover, \$4.95

The Struggle to Bring Technology to Cities, 1971, URI 70001, Paperback, \$1.95

How Clean is Our City? A Guide for Measuring the Effectiveness of Solid Waste Collection Activities, Louis H. Blair and Alfred I. Schwartz, 1972, URI 10004, Paperback, \$1.95

How Shall We Collect The Garbage? A Study in Economic Organization, Dennis R. Young, 1972, URI 10008, Paperback, \$1.95

Measuring the Effectiveness of Local Government Services: Recreation, Harry P. Hatry and Diana R. Dunn, 1971, URI 70002, Paperback, \$1.75

An Introduction to Sample Surveys for Government Managers, Carol H. Weiss and Harry P. Hatry, 1971, URI 30003, Paperback, \$1.50

Obtaining Citizen Feedback: The Application of Citizen Surveys to Local Governments, Kenneth Webb and Harry P. Hatry, 1973, URI 18000, Paperback, \$1.95

The High Cost of Education in Cities: An Analysis of the Purchasing Power of the Educational Dollar, Betsy Levin, Thomas Muller, and Corazon Sandoval, 1973, URI 31000, Paperback, \$2.50

University Urban Research Centers, Second Edition, Grace M. Taher, Editor, 1971-72, URI 10002, Paperback, \$2.75

Cable Television in the Cities: Community Control, Public Access and Minority Ownership, Charles Tate, Editor, 1971, URI 80004, Paperback, \$3.95

Governing Metropolitan Areas: A Critical Review of Councils of Government and the Federal Role, Melvin B. Mogulof, 1971, URI 70004, Paperback, \$2.25

Measuring the Effectiveness of Local Government Services: Transportation, Richard E. Winnie and Harry P. Hatry, 1972, URI 16000, Paperback, \$1.95

**COASTAL ZONE
INFORMATION CENTER**



THE URBAN INSTITUTE

2100 M Street

NOAA COASTAL SERVICES CENTER LIBRARY

0037



3 6668 00002 5132