DOCUMENT RESUME

ED 183 614 TM 800 023

AUTHOR Schmidt, Richard E.; And Others
TITLE Evaluability Assessment: Making Public Programs Work Better. Human Services Monograph Series. Number 14.
INSTITUTION Department of Health, Education, and Welfare, Washington, D.C. Project Share.
REPORT NO DHEW-OS-76-130
PUB DATE Nov 79
NOTE 110p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Administrative Personnel; Data Collection; Decision Making; Expectation; *Federal Programs; Information Utilization; Measurement; Models; Policy; *Program Descriptions; *Program Evaluation; *Program Improvement; Reliability
IDENTIFIERS *Evaluability Assessment; Urban Institute

ABSTRACT

Evaluability assessment is a descriptive and analytic process intended to produce a reasoned basis for proceeding with an evaluation of use to both management and policymakers. It was jointly developed by the members of the program evaluation group of the Urban Institute between 1968 and 1978. The approach begins by obtaining management's description of the program. The description is then analyzed to determine whether it is: complete, acceptable to the policymakers, and a valid representation of the program as it actually exists. Three types of models are used: logic models, function models, and measurement models. The evaluator determines whether management's expectations are plausible, whether its evidence is reliable, and whether the program is feasible to develop. Perhaps the most common problem affecting federal programs is unrealistic expectations, most likely resulting from a management decision which is invalid, implausible, or unacceptable to policymakers. A task for the evaluator is to construct alternatives wherever possible. The final step is to implement evaluability assessment findings. (Author/CTM)

Evaluability Assessment: Making Public Programs Work Better

PROJECT SHARE

A National Clearinghouse for Improving the Management of Human Services

Human Services

NUMBER 14 NOVEMBER 1979 MONOGRAPH SERIES

Evaluability Assessment: Making Public Programs Work Better

RICHARD E. SCHMIDT
JOHN W. SCANLON
JAMES B. BELL

PROJECT SHARE

A National Clearinghouse
for Improving the Management
of Human Services

PROJECT SHARE has contracted for the preparation of a monograph series in order to survey the state of the knowledge or the state of the literature in selected subject areas of importance to the human services community. The monograph series will also provide an opportunity for authors to offer their views and opinions on these topics. It is the aim of PROJECT SHARE to stimulate discussion through the publication of these monographs.

The views and opinions expressed in this monograph are entirely those of the authors and are not necessarily those of the Department of Health, Education and Welfare, PROJECT SHARE or Aspen Systems Corporation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This manual was completed under a contract with the Department of Health, Education, and Welfare. Much of the work and many of the examples are thus taken from the Bureau of Health Planning. The concept of evaluability assessment, however, has been in the process of development for at least 5 years. We drew upon that entire history to write this manual. Where appropriate, we have included examples which are more completely documented in separate reports from other agencies. Several instances are cited in which the authors, their colleagues, and our work would have benefited from evaluability assessment. We believe that this manual represents a good beginning definition of what we hope to be a promising approach to management and evaluation of Government programs. We hope that it will continue to be refined by other practitioners.

Although the context for the process described here is Federal, and most references are to Federal managers, the concepts which we outline are applicable to any goal-oriented organization responsible for managing public programs.

Our intent in developing this method is to provide techniques that can be used in making public programs work better than they now do. We have aimed the document at evaluators who are charged with informing management and policy officials about the effectiveness of their programs. The method, however, goes beyond providing information. We intend for it to forge closer and more useful links between management and evaluators, hopeful that the net result will be better program management and more effective programs.

ACKNOWLEDGMENTS

The authors completed this work under a contract with the Department of Health, Education, and Welfare (DHEW). We are indebted to the many individuals within the Department for the time and effort they devoted to the development of the evaluability assessment method. Robert Raymond of DHEW deserves special mention because it was largely due to his interest and perseverance that we were able to test the concept. The management and staff of the Bureau of Health Planning contributed generously of their time during the evaluability assessment of Public Law 93-641. We would especially like to thank Dr. Harry Cain, the former director; Dr. Colin Rorrie, the current director; Messrs. Samuel Stiles, Jeff Human, Tom Croft, and Roland Peterson, division managers; Ms. Judy Lewis and Mr. Jon Gold, key staff from the Bureau's evaluation office; and Mr. William Berry, acting director, Bureau of Health Facilities.

This guideline, although flowing directly out of our work with DHEW, is also based on prior work completed by our colleagues at The Urban Institute's program evaluation project. Between 1968 and 1978, the following professionals contributed to the effort.

Deborah Anderson
Martin Berlin
Peter Bloch
Joan Brackett
Harrington Bryce
Garth Buchanan
Ilona Bush
Sumner Clarren
Hugh Duffy
Thomas Fischgrun
William Foskett
Warren Frederick
James Fukumoto
Pamela Gervais
Lucile Graham
John Haberle
John Heinberg
Eric Hollins
Pamela Horst
Barbara House
Peg Kay
Sara Kelly
Dona Kemp
Sharon Kirby
David Klaus
Cynthia Lancer
Timothy Ling
Reynold Madoo
Michael Mulkey
Paul Nalley
Joe Nay
Montina Pyndell
Elliot Ratner
Katryna Regan
Alfred Schwartz
Ralph Schwartz
William Tash
Francine Tolson
John Trutko
Cyrus Ulberg
Alease Vaughn
Leona Vogt
John Waller
Gilbert Ware
Steve Watson
Bayla White
Thomas White
Thomas Wolfsohn
Jane Woodward
Richard Zamoff

In particular, Dr. Joseph Wholey directed the evaluability assessments completed for the Bureau, thus contributing much of the substance of this manual. We would like to thank Dr. Donald Weidman of the Office of Management and Budget for his careful and helpful review of the manuscript and for his contribution to the development of the method while he was an associate at The Urban Institute.

Over the years, the following organizations supported development of evaluability assessment:

- Department of Labor
- Department of Justice, Law Enforcement Assistance Administration
- Police Foundation
- Department of Health, Education, and Welfare
- Department of Commerce

- General Accounting Office
- Appalachian Regional Commission
- Legal Services Corporation of America
- New England Regional Commission
- Department of Housing and Urban Development
- Department of Energy
- And the many regional, State, and local agencies who have helped by cooperating with our effort, supplying raw data, and reacting openly to our findings.

Finally, we owe a debt of gratitude to Ms. Elaine Moses, Ms. Mary Sarley, Ms. Elaine Turkas, and Ms. Chris Harding who typed and retyped the several drafts of the manuscript.

To all those who helped us, we say thank you.

I. EVALUABILITY ASSESSMENT: WHY IS IT NEEDED? WHAT IS IT?

The Problem and a Solution

The ritual enactment of legislation which provides for publicly financed programs engages the executive and legislative branches of government in a compact—an agreement on the existence of a problem and of a solution to it. Even if the players believe the solution to be a partial one, they hold high hopes for it. Both sides share a dream, even if only momentarily, that this program will work to produce the benefits intended. They assume that agency and program managers will implement their solution wisely and manage the resources placed at their disposal to achieve the desired ends. They foresee oversight hearings in which objective evidence of performance is provided by agency executives, enabling legislators then to modify the legislation accordingly. Finally, they hope, the program will be judged by the public and their elected representatives as a good investment, ably managed. One small piece of this dream is the assumption that program managers can and will convert the legislation into a workable program, and that they will have the necessary information to help them direct the program to the desired end.

To this end, Congress and the executive branch have increasingly invested in program evaluation over the past decade. Starting from nearly nothing in the early sixties, investment in evaluation grew to around a quarter of a billion dollars by 1976. Unfortunately, however, the investment has not yet paid off. Program evaluation has not led to successful policies or programs. Instead, it has been planned and implemented in isolation from Federal decisionmaking, and has produced little information of interest and utility to policymakers and managers.

If Federal policies and programs were always successful, then the limited utility of the evaluation product would be a minor irritant. However, not only are few programs demonstrably successful, but many others are viewed by the public and elected officials as ineffective, wasteful, and even harmful. Since programs are usually a response to some perceived need or problem, abandoning them is not a politically viable solution. *The only solution is to make programs work and be able to demonstrate that they do.* Indeed, program evaluation was created to help improve policies and programs by measuring actual performance and by sifting through what works well and what does not.

Why has practice fallen so short of expectations? Were the expectations held out for evaluation unrealistic? Is it true, as many observers feel, that blame for the failures of evaluation should fall on Congress? For 10 years the literature has called for more precise objective definition of what needs to be evaluated to be included in legislation. Evaluation would then be able to produce the answers so desperately needed. This viewpoint, however, is both incorrect and dangerous.

Congress does not have and cannot get enough information to set measurable objectives which are also realistic, plausible, and achievable. It is reasonable only to expect the Legislature to set goals at a conceptual level (i.e., "put a man on the moon within a decade") and then watch to insure that program managers are translating objectives into realistic and acceptable measurable definitions of congressional intent. It is the program manager who should develop the information required to set such objectives, and the program evaluation function is one important source of that information.

Yet as many have noted, the program evaluation functions established in government have not been used by program managers. The authors of this manual believe that the failure of evaluation to help produce effective policies and programs stems from a number of factors relating both to the way programs and the function of evaluation are designed and managed.

In our research over the last 5 years, we have defined a number of conditions (Horst et al., 1974) which, if present, generally disable attempts to evaluate performance. More fundamentally, though, these conditions also disable attempts to manage programs for important results of national concern. Having identified such roadblocks, we attempted to develop methods to eliminate them. The result of that effort is a method we call evaluability assessment (EA).

Evaluability assessment is an evaluation technique and management decision process designed to deal with the recurring problem of managers and policymakers who frequently are unable to guide programs toward desired performance and outcomes. Even with substantial evaluation efforts, some programs do not improve over time. We believe this is because of deficiencies in the descriptions of programs and the administrative structure of government that hamper management and evaluators. With EA, the manager is helped to establish conditions that are prerequisite to success.

Evaluability assessment begins with an analysis of management's description of its program to determine whether or not it meets the method criteria. When even one of the questions is answered "no," there is a high probability that the program will not be successful and evaluation information will not be useful. The existing program description is said to be *unevaluable* since no useful evaluation is likely.

Part of EA requires the evaluator and management to work together to remove conditions that make the program description unevaluable. The result of an EA is a set of management decisions that establish an evaluable program description. With such a description management can be confident that evaluation will show the program to be successful in terms acceptable to policymakers and will provide management with the capability to achieve performance.

What Is Evaluability Assessment?

Evaluability assessment is a descriptive and analytic process intended to produce a reasoned basis for proceeding with an evaluation of use to both management and policymakers. It was jointly developed by the members of the program evaluation group of The Urban Institute between 1968 and 1978. While all versions of the method are based on the applied experiences of the members of the group, definitions and sequence of operations vary slightly among the different practitioners. Joseph S. Wholey, in *Evaluation: Promise and Performance* (in press), describes his approach.

Basically, the group has taken the problems we have observed over the past 5 years and converted them into a set of criteria which must be satisfied before proceeding to full evaluation. The criteria are used to guide the evaluation through a series of planning and analysis tasks designed to identify and then eliminate roadblocks.

The approach basically begins by obtaining management's description of the program. The description is then systematically analyzed to determined whether it meets the following requirements:

- It is complete.
- It is acceptable to policymakers;
- It is a valid representation of the program as it actually exists;

- The expectations for the program are plausible;
- The evidence required by management can be reliably produced;
- The evidence required by management is feasible to collect; and
- Management's intended use of the information can realistically be expected to affect performance.

Each of these points is a question to be answered by the evaluator, working with the manager. We also add a function for the evaluator which is not traditional—when the analysis reveals gaps or problems in management's description, he should be involved in devising alternative descriptions which will resolve the problems. This function puts the evaluator into both program and management system design, based on an assumption that if evaluation is to be used by management to affect performance, it is inseparable from these design needs.

What Are the Conclusions an Evaluability Assessment Can Reach?

The object of an EA is to arrive at a program description that is evaluable. In other words, the description permits the program to be measured with some reasonable assurance that the evaluation can be done and that predetermined expectations can be realized. As noted previously, if even one criterion is not met, the program is judged to be unevaluable, meaning that there is a high risk that management will not be able to demonstrate or achieve program success in terms acceptable to policymakers

There are three types of unevaluable conclusions that can be reached:

1. *Comparison of program management's description with the list of prescribed elements show the description is incomplete.* Certain parts of the description are not available, indicating that management has not yet made some key decisions. For example, an objective may not have been defined in terms of measures and expectations. Usually there are people with ideas about how to define an objective, but the management decision on definition may be missing. An incomplete description can be easily fixed by management.

2. *Comparison of the evaluator's and operator's description with program management's description indicates the program management's description is implausible, invalid, not cost-feasible to evaluate, or not useful* Problems raised by this comparison mean that either management's description is inadequate or the evaluator's description is wrong. The evaluator will present evidence to show why management's description is unevaluable. If management disagrees, the evaluator then requires additional information on the program activity to resolve the disagreement.

3. *Comparison of policymakers' and program management's description shows that program management's description is unacceptable* If policymakers disagree with program management, it is generally because program management has either omitted an objective or set of activities important to policymakers or included objectives or activities with which policymakers disagree. What is needed in such cases is a process whereby disagreements can be aired and validated and missing elements added through a program design activity introduced by management. It may be possible to change policymaker descriptions after analysis of the areas of disagreement.

An evaluable conclusion, on the other hand, produces a set of decisions made by management regarding the following areas.

(1) evaluable program description;
(2) evaluation information to purchase;
(3) whether to change the design of the program;
(4) whether to change the expectations of policymakers;
(5) how to organize and staff implementation of the decisions.

Managing an Evaluability Assessment

This manual describes the series of activities which we believe to be necessary in completing an evaluability assessment. Two things must be said about managing such an enterprise. First, the sequence of such activities described here should not be taken as the only proper one. While it is logical and orderly, it is sometimes made impossible by the reality of events. Much of the information needed must be obtained from policymakers, managers, and program operators, all of whom are busy people, and it is usually up to the evaluator to change his schedule to suit theirs. Second, many steps or analyses defined separately here are often done together, and as often rather more informally (some might say haphazardly) than is implied here. Even using our method of analyzing programs prior to evaluation, no two completed assessments look alike.

We expect that many evaluators who attempt to follow this manual will discover that our rules and procedures are imperfect, and we hope that new or different procedures will be developed by those who attempt to use the method. Still, it must be remembered that often the evaluation team will have a fixed budget and schedule to complete the EA. Within these constraints, the project will have to adjust resources and products to produce an assessment that serves management best, a judgment that the evaluator should make jointly with program management. Immediately following are some guides to planning and controlling the work, and subsequently, the reader will find some specific how-to guidance.

Planning an Evaluability Assessment

The evaluation team will spend most of its time in the following activities:

- Managing the EA: planning the work, scheduling activities, controlling and adjusting staff allocation to activities, monitoring and controlling the quality of the work;

- Reviewing the literature: reading and becoming familiar with the legislation, hearing records, program plans, guidelines and regulations, research and evaluation reports, documentation of field activity, etc.;

- Interviewing policymakers (includes upper level bureaucrats; Office of Management and Budget staffers, Congressional aides of Senate and House authorization and appropriations committees), program management, and a sample of regional office staff and State staff (when appropriate);

- Documenting program descriptions obtained from literature search and interviews and communicating with key officials to verify the accuracy of the descriptions.

- Conducting site visits to observe program activity for the purpose of describing direct and indirect intervention;

- Completing analyses to determine evaluability;

- Documenting (in briefings or written form) the conclusions and supporting evidence; and
- Organizing and administering the decision process following from the assessment.

During this period, management is expected to participate also. Their activity will include:

- Providing program description to evaluator;
- Assessing the validity of the evaluator's descriptions and conclusions;
- Providing guidance to the evaluator on the relevance and direction of the work; and
- Making decisions on options developed by the evaluator to resolve problems affecting evaluability.

Table 1 indicates roughly the level of effort that might be expected for these activites. The budget for the evaluation team in this case is estimated at 8 person-months; the time available, 4 calendar months. During this time, the evaluation team must stay in contact with management, providing information as the effort proceeds.

Controlling the Work

Controlling the work is perhaps no different from controlling any other analytic effort, all of which have a way of expanding as the work progresses. The team leader will have to judge whether a particular activity is beginning either to exceed the expected value or diverge from the main path. A semblance of control is provided by tracking the consumption of man-hours on each activity and tracking schedules As these are flexible, however, a team leader cannot expect too rigid an adherence to original estimates—except that maintenance of overall cost and schedule constraints is important

We assume that more than one person will be involved in any EA, and have, in fact, found the optimum team size to be two or three people. One obvious advantage of such a team is that if one or more members become stuck on some issue or problem, the group can jointly decide what to do about it Often we have found ourselves stuck on some point and discovered either we needed more specific data or were trying to make some decision that was not ours to make. As a group, we were able to determine that contact with program management was necessary.

Finally control is afforded by frequent progress checks against a checklist of activities and products Such reliance on the list is important, simply because it is so easy to drift away on interesting questions or a quest for more conclusive data The evaluator also must always bear in mind that the effort is *not* a complete evaluation study—rather it is an evaluability assessment. He must do only enough to allow himself to move to the next stage. While the potential data and analytic questions are limitless, the time and resources, as well as the patience of program management are not

TABLE 1 Time Spent on EA Activities in a Four Calendar-Month Period

Activities	Evaluators' Time (Person-Weeks)
Managing the EA	2
Reviewing Literature	4
Interviewing policymakers	4
Documenting Program Descriptions	4
Visiting Sites	4
Completing Analyses	6
Documenting Findings	4
Organizing the Decision Process	4
TOTAL	32

Activities	Management's Time (Person-Weeks/Person)
Providing Information	2/5
Judging Validity	1/10
Providing Guidance	1/10
Making Decisions	2/5
Total	1

II. THE FIRST STEP: DESCRIBE THE PROGRAM

Evaluability assessment is a 4-step process: (1) describe the program; (2) validate and analyze the program description to determine the extent to which the program can be evaluated; (3) construct alternatives wherever necessary; and (4) implement management decisions.

This first step is crucial in evaluation planning and program management. There is often a substantial discrepancy between the concept of a program as perceived by managers and policymakers and the reality of that program as it exists in the field, as well as a hesitancy on the part of management to admit to the discrepancy. Part of the problem is no doubt a simple unwillingness to admit that management at the Federal level does not and probably cannot control what happens at the local level in most programs. Acceptance of this truth may seem to threaten their positions and seriously weaken their support in Congress.

Yet continued unwillingness to accept the differences between the reality and the idealized perceptions of those programs can seriously limit the ability of management to obtain useful performance information and influence or control the events and outcomes. It is for this reason that we have developed an approach to evaluation planning that provides a method to describe programs and determine the extent to which such discrepancies exist. We do this within the context of program management in as nonthreatening a manner as is possible. The method is intended to help program managers gain whatever control is possible—to help them achieve their ends, as well as satisfy the demands of congressional and executive branch oversight functions.

What Is a Program Description for Purposes of EA?

The core of an evaluability assessment is program management's description of the program. This involves two elements—real programs and various descriptions of real programs. Descriptions are almost invariably highly condensed summaries of real programs, which nearly never capture all activities and effects involved in a program. This lack is exaggerated by the fact that program descriptions are often inaccurate, reflecting what people think, want to believe, or want others to believe is actually happening. To the extent they do not incorporate key program characteristics necessary for evaluation or are inaccurate, we believe that evaluations based on those descriptions will be equally inaccurate. Thus, in evaluability assessment, we continually check various program descriptions against one another and against the main referent—the program as it is observed to actually operate.

What Is Being Described in an EA?

For purposes of an EA, the program description should cover those activities which management can guide and control to achieve national objectives. Those activities can be broken out as:

- *Direct Intervention*—activities designed to directly change or serve society in some desired way;
- *Indirect Intervention*—activities through which management acts to develop and alter the direct intervention and its performance;
- *Program Accountability*—activities through which management acts to report performance and establish realistic expectations among policymakers.

The components of each are illustrated in Figure 1.

The *direct intervention* is usually at the interface between government and the recipient of the government service. Examples include community mental health centers and their clients; Head Start projects and their pupils; Occupational Safety and Health Administration regulations, inspectors, and the workplace; and the U.S. employment service offices, the job seeker, and the employer.

Indirect interventions are a type of overhead function whose purpose is to assure and improve the performance of the direct interventions. Evaluation systems, technical assistance programs, regional office monitors, State and local planning agencies, research and demonstration programs are all examples of indirect interventions. The term "indirect" is used because the indirect intervention must first change the behavior of some other organizational unit before program performance can be altered.

Program accountability can occur through formal decision processes such as budget and oversight procedures or through informal mechanisms such as publications, meetings, correspondence, and conferences. Federal program offices establish staff activities to deal with both the formal and informal channels of communication with the mission of enabling management to get agreement with policymakers on realistic expectations for the program and informing policymakers on progress and program performance.

What Comprises a Program Description?

Programs can be described in many ways, but for purposes of an EA certain specific types of information should be included. Our scheme for describing programs consists of elements that constitute a description and properties or criteria that each must satisfy to be judged evaluable. If one or more elements are missing or one or more criteria are not met, the probability increases that evaluation will be useless and measured program performance unacceptable.

Elements The following elements are required for a description (see Table 2)

- a *sequence of events*;
- an *event description*;
- *measures* and *comparisons* defining each event activity.
- *expected values* for each event;
- *activities* that must exist for event to occur;
- *information system* used to provide evidence the event occurs; and
- *intended use* of evaluation information.

Event sequences are used to define the ordered chronological interrelationships among events. In part the sequence defines the logic—the assumed cause and effect relationships among events. It also implies an order in which measurements could be taken. Finally, sequences suggest possible stopping points in measurement; if we know that Event A has not occurred, we may not need to measure Event B.

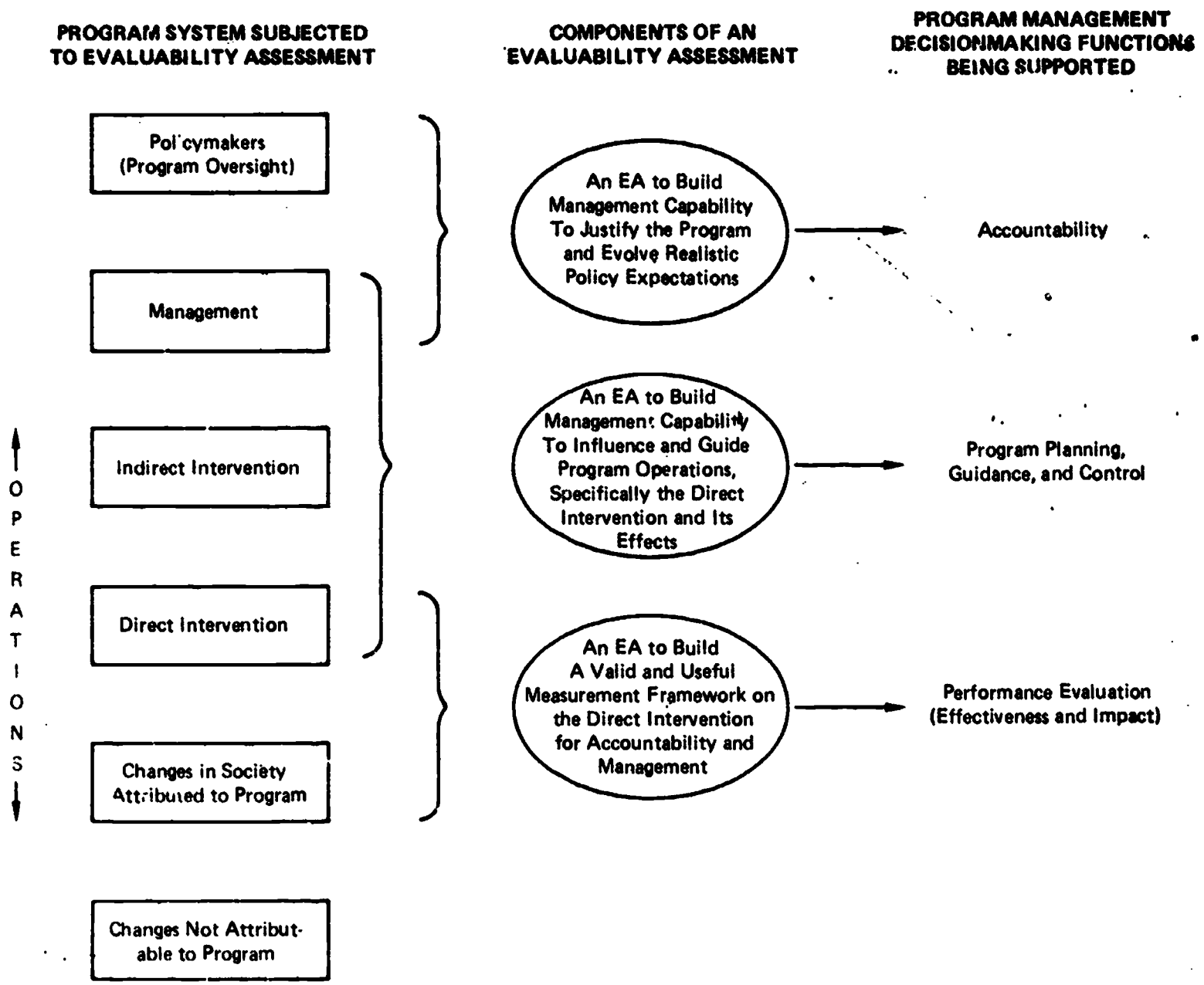


FIGURE 1 Distribution of the Tasks in an Evaluability Assessment

TABLE 2 Elements Needed in a Complete Program Description

Element of the Program Description	1. Event Sequence That Describes Program Behavior	2. Measures and Comparisons Describing Each Event	3. Expected Values For Each Event	4. Activities That Must Be in Place For Event To Occur	5. Information Systems Used to Provide Evidence of Event Occurring	6. Use of Evaluation Information
Information items that must be known for the program description to be well defined	1. Sequence begins with events over which management has direct control 2. Sequence includes all events that must occur for objectives to be achieved 3. Events include achievement of all objectives considered necessary to justify the program 4. Events include all positive and negative expected side effects	1. A set of measures and comparisons considered necessary and sufficient by owner to describe the event is identified 2. Measures and comparisons also describe the evidence acceptable to the owner demonstrating that the event occurred	1. The owner's* expected values for each event are identified 2. Time periods in which events should occur are given	1. The program activities that must be in place for each event to occur are identified 2. The characteristics of those activities necessary for expectations to be met are specified 3. The rationale or evidence indicating that the activity will achieve expected results is identified	1. The measurement systems that provide data on each event are identified 2. The measurement operation and instrument is described for each measure 3. Methods and mechanisms available to estimate and check for measurement and processing errors are identified 4. Costs of measurement systems are known	1. The users of evaluative information are identified for each event 2. Actions or processes for defining action to be taken on evaluative information are identified 3. How these actions are expected to affect the program or policies described

*"owner" refers to person or group from whom information is obtained, e.g., program manager.

Event descriptions are used to define the event, as in "children from X-type families will read as well as children from Y-type families." Attempts should be made to avoid event descriptions which include or imply two events, as in "children from X-type families will read as well as children from Y-type families as a result of adopting Z-type curriculum." Such events should be described in at least two statements.

Measures and *comparisons* are used to define the type of evidence that will signal event occurrence. In the event above, the measure might be a comparison of performance on a standardized reading test in children from X-type families with children from Y-type families. Comparisons can be one group against another, one group against itself over time, or both.

Expected values are used to define the direction and amount of change expected. If we say, reading scores will increase " then implicitly any increase is acceptable. A better statement would be that reading scores will increase 25 percent in one year or that scores between two currently unequal groups will become equivalent. Often it is useful to define any expected change in a comparison group to avoid a situation in which, for example, reading scores become equal because of a reduction in comparison group scores.

Activities are used to define and help in examining the plausibility of the basic program design. Subsequently, measures of the types and amount of activity can be used to help explain performance.

Information systems are used to define the sources which will be used to collect data, especially as to whether the sources now exist or new systems will need to be created.

Intended use helps to define and examine the probability that evaluation information will be worth its cost.

Element Properties. The event itself must begin with an activity over which management has control, conclude with activities representing the primary intent or objectives of the program, and include all of the key events necessary for the objectives to be achieved. Selection of key events is judgmental, so evaluators and others must be selective in defining the necessary and sufficient conditions for the objectives to be achieved.

In addition to the properties, EA requires an analysis of the program description to assess the extent to which the description satisfies evaluability criteria based on eliminating typical problems which have prevented or inhibited useful evaluation in the past. The following standards must be met:

- *Well defined*—the description is well defined when it satisfies the requirements of Table 2;
- *Acceptable*—the description is acceptable when it matches policymaker expectations;
- *Valid*—the description is valid when it accurately represents the program activities actually operated in the field;
- *Plausible*—the description is plausible when there is evidence of plausibility and no evidence of implausibility;
- *Data feasibility*—the data system defined in the description is feasible when it does not impose cost or political burdens beyond those which management/policymakers are prepared to accept;
- *Data reliability*—the data system defined in the description is reliable when provisions are built in for repeat observations and further verification prior to use;
- *Plausible use of information*—management's expectations for the program and the evaluation are plausible when its use of the information is defined in terms of known resources and activities within its control.

These criteria are elaborated in terms of rules and procedures in Chapter III. Table 3 illustrates the relationship between these EA criteria and the elements of a description.

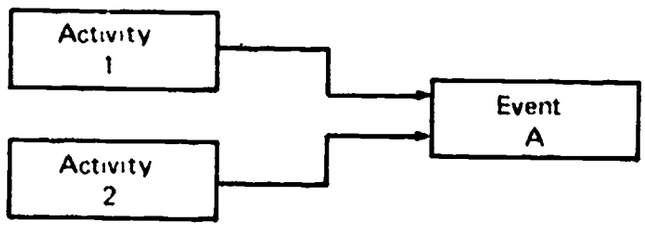
Models: A Useful Language Convention in Describing and Analyzing Programs

We use three types of models to describe programs:

- Logic models: diagrams representing the intended logic of the program, i.e.,

If Event A occurs, then it is assumed Event B will occur, with events representing program effects or objectives.

- Function models: diagrams representing the intended activities or functions of the program which support, produce, or lead to the intended events, i.e.,



- Measurement models: diagrams representing the points at which measures can be taken, as well as the types of measures needed to represent activities and events, i.e.,

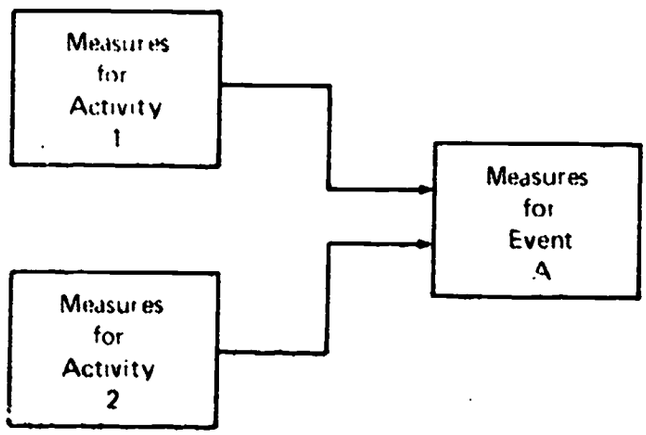


TABLE 3 Evaluability Assessment Criteria

Element	Criteria
Event Sequence	• Acceptable • Well Defined
Event Description	• Acceptable • Well Defined • Valid
Measures	• Acceptable • Well Defined
Expected Values	• Acceptable • Well Defined • Plausible
Evidence	• Acceptable • Well Defined • Cost-Feasible • Known to be Reliable
Use of Information by Management	• Acceptable • Well Defined • Plausible

Each of the models can be used to establish the extent to which the program description satisfies the EA criteria (Table 4).

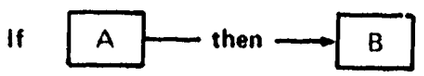
TABLE 4 Use of Models in EA

EA Criteria	Logic Model	Function Model	Measurement Model
Well Defined	X	X	X
Acceptable	X	X	
Valid	X	X	X
Plausible	X	X	
Feasible Data			X
Reliable Data			X
Plausible Use of Information		X	X

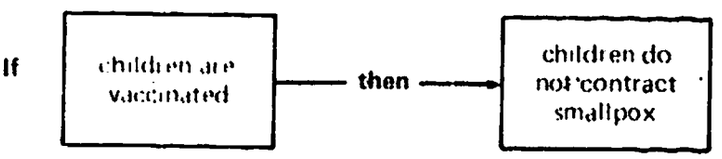
It should, of course, be remembered that other types of descriptions are possible, although we advocate development and use of these models in an EA. The key issue is not the method used, but the information it elicits, as well as the adoption and use of the EA criteria.

Logic Models

We noted that logic models are of the type:



A simple example would be:



or

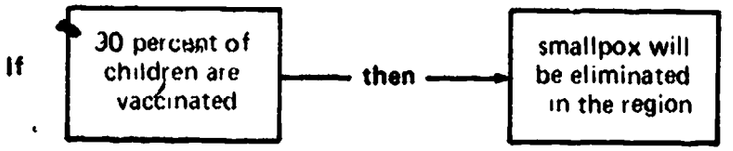


Figure 2 illustrates a logic model (Schmidt et al., 1976) developed to describe the health and child development program operated by the Appalachian Regional Commission. The model represents the intended events connected with one part of the overall program—health and child development—and was used to define this specific program prior to the decision regarding an approach to evaluation.

Function Models

Function models are intended to define the program in terms of specific activities and flows of people, money, and information. Often called process or stock and flow models, they are intended to capture the operations or processes which characterize a program or project.

Figure 3 illustrates a function model developed to define one part of a demonstration program (Horst. 1978) funded by the Law Enforcement Assistance Administration (LEAA)—Managing Criminal Investigations (MCI). The function represented is case processing, the flow is information.

Measurement Models

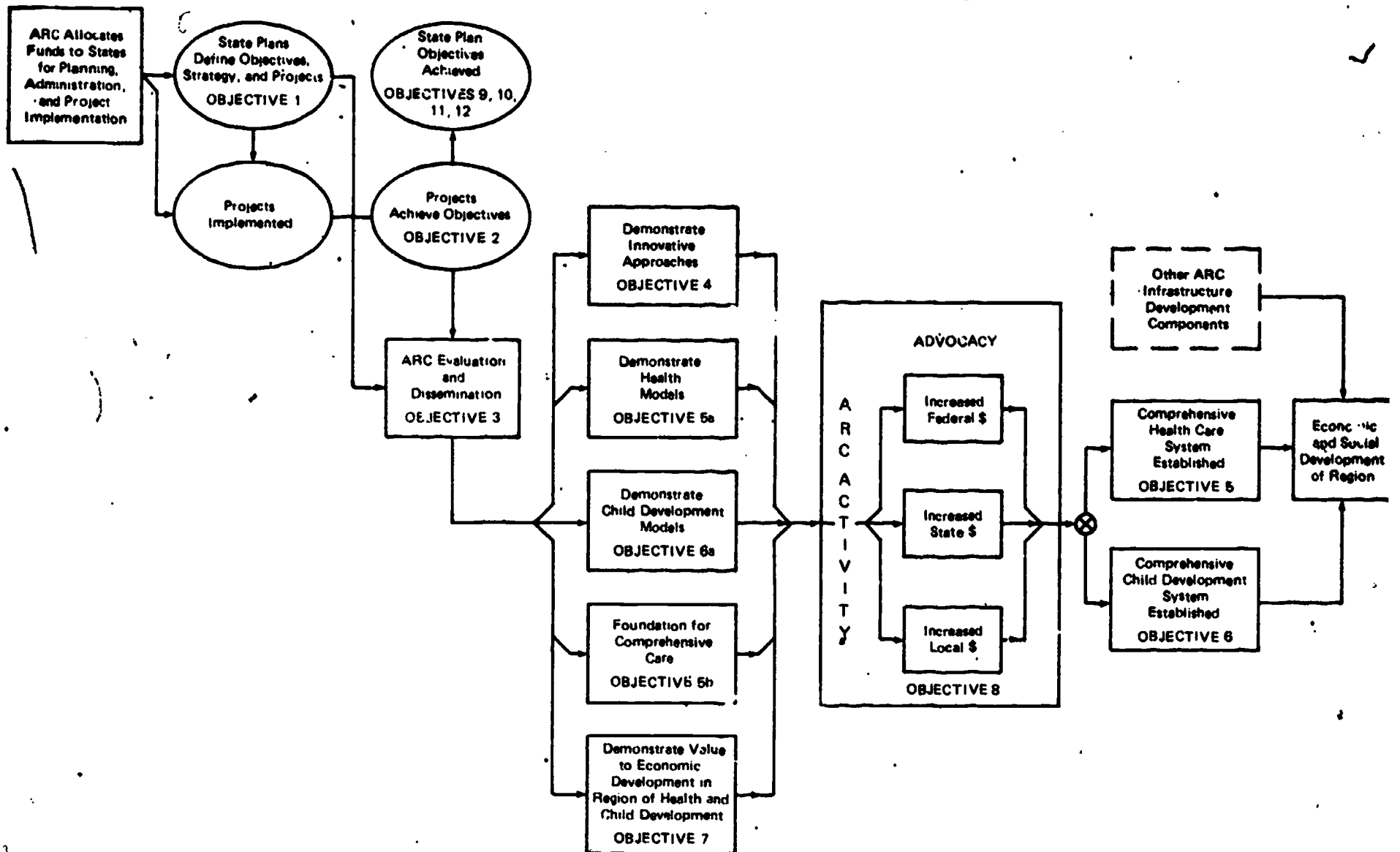
Logic and function models can be used by themselves for defining, analyzing, and comprehending a program of some complexity, but measurement models can only be used in conjunction with the others. To illustrate the interrelationships of the models, we have excerpted material from a report completed for LEAA (Nay et al., 1977-Annex D) concerning the use of measurement models. Figure 4 illustrates a function model for a city methadone treatment program by showing people flowing through various functional activities (e.g., in supervised methadone treatment, event 6).

In the accompanying table, each state or condition of the population involved is defined, the measures are identified, and current availability of a data collection system is noted. This model is a relatively simplified representation of such a program.

Figure 5 then represents the logic and measurement models for this program. The events and assumptions at the top represent the program logic. The evaluation questions and measurements and comparison rows represent questions of interest and measurements to be taken from the function model. Note that the numbers in the measurements and comparisons entries represent measurement points taken from the function model, Figure 4, and that some of the questions ask for descriptions of how whether certain events occur. Note, too, that existing data systems are inadequate to answer all questions.

In conclusion, it must be again stressed that the models are nothing more than a language used to define certain aspects of a program. We find models useful for our analyses, but other forms have been and are used to define programs.

FIGURE 2 Appalachian Regional Committee Health and Child Development Program Model



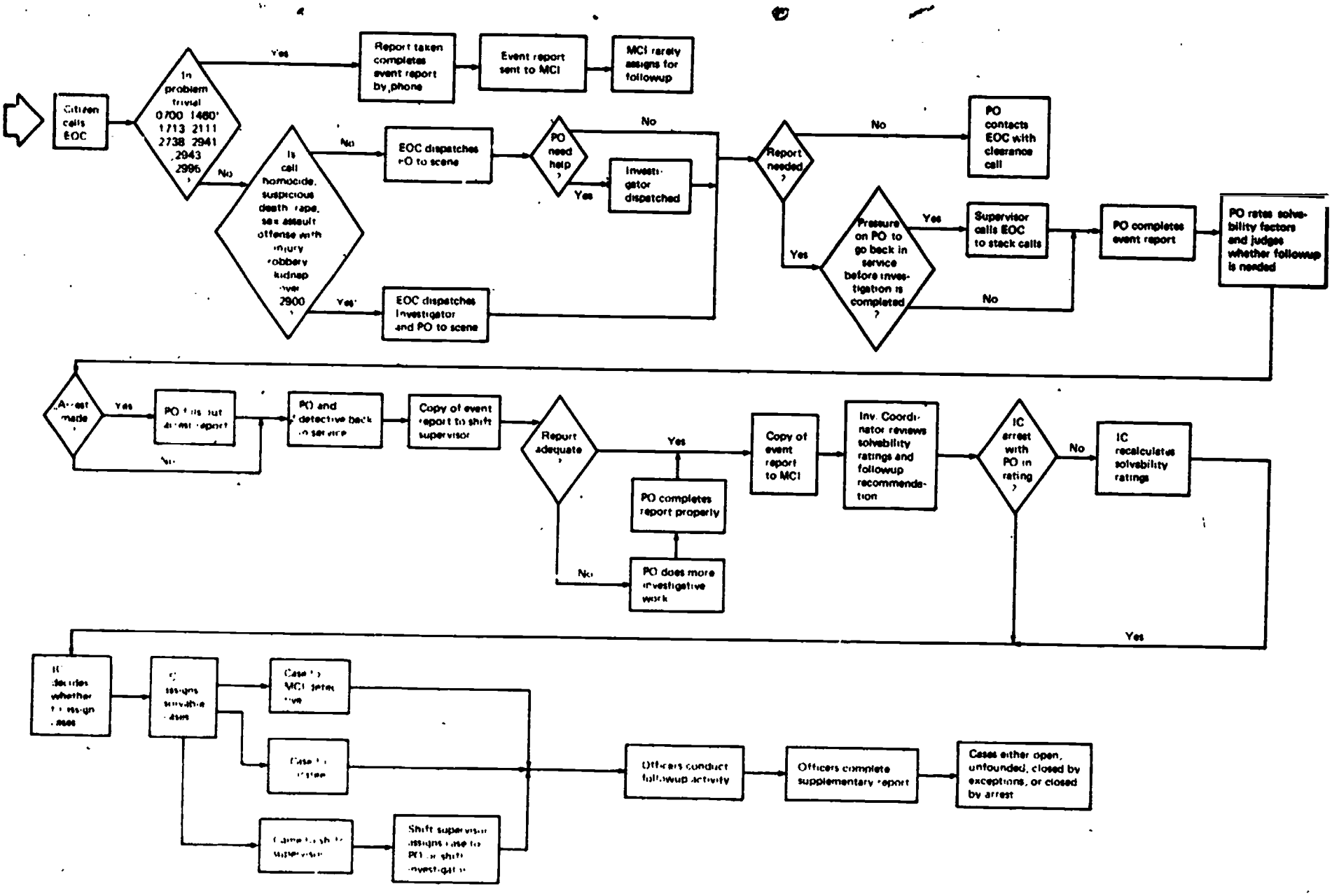


FIGURE 3 Case Processing Function Model Emphasizing Patrol-Related Activity for Managing Criminal Investigations (MCI)

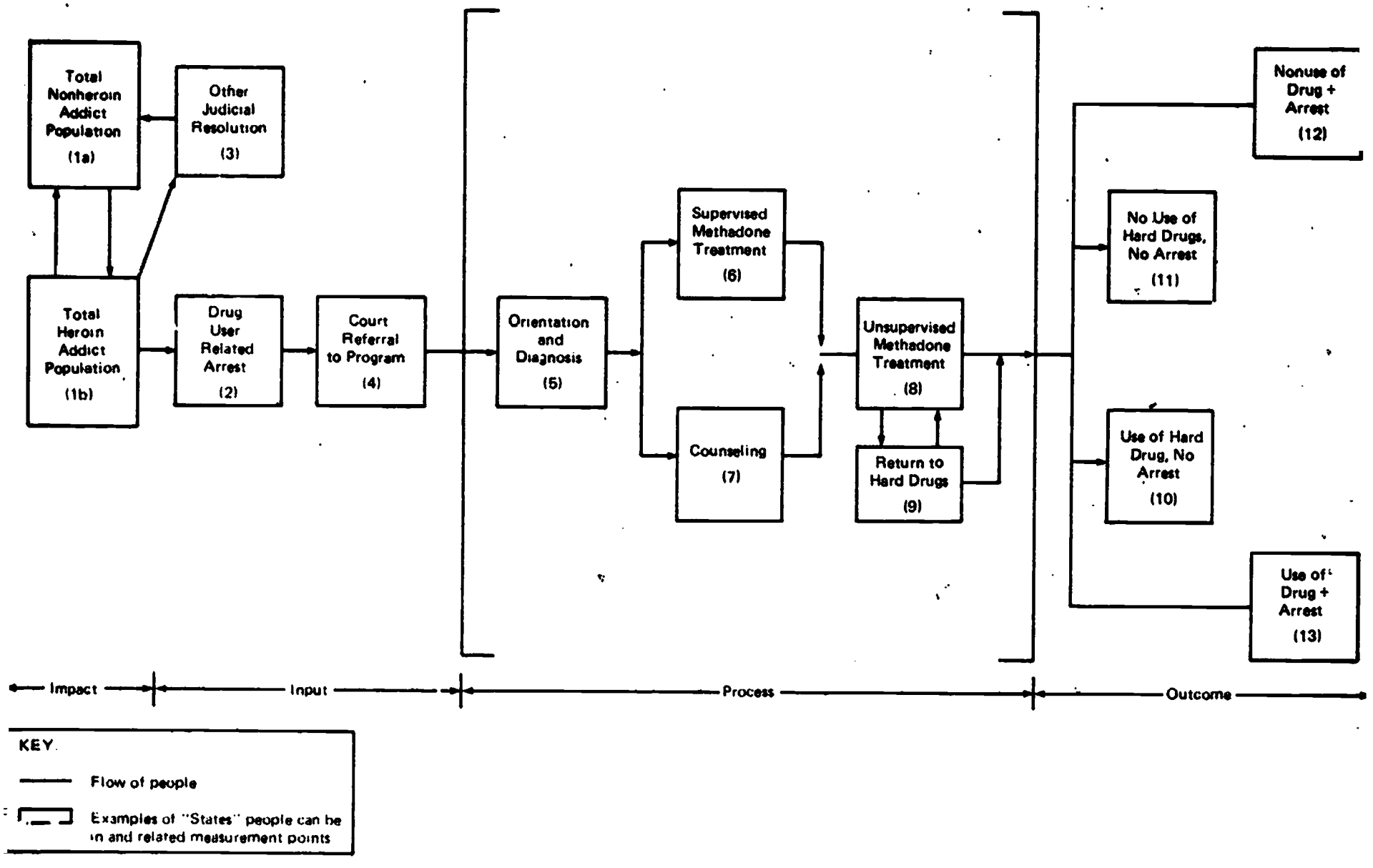
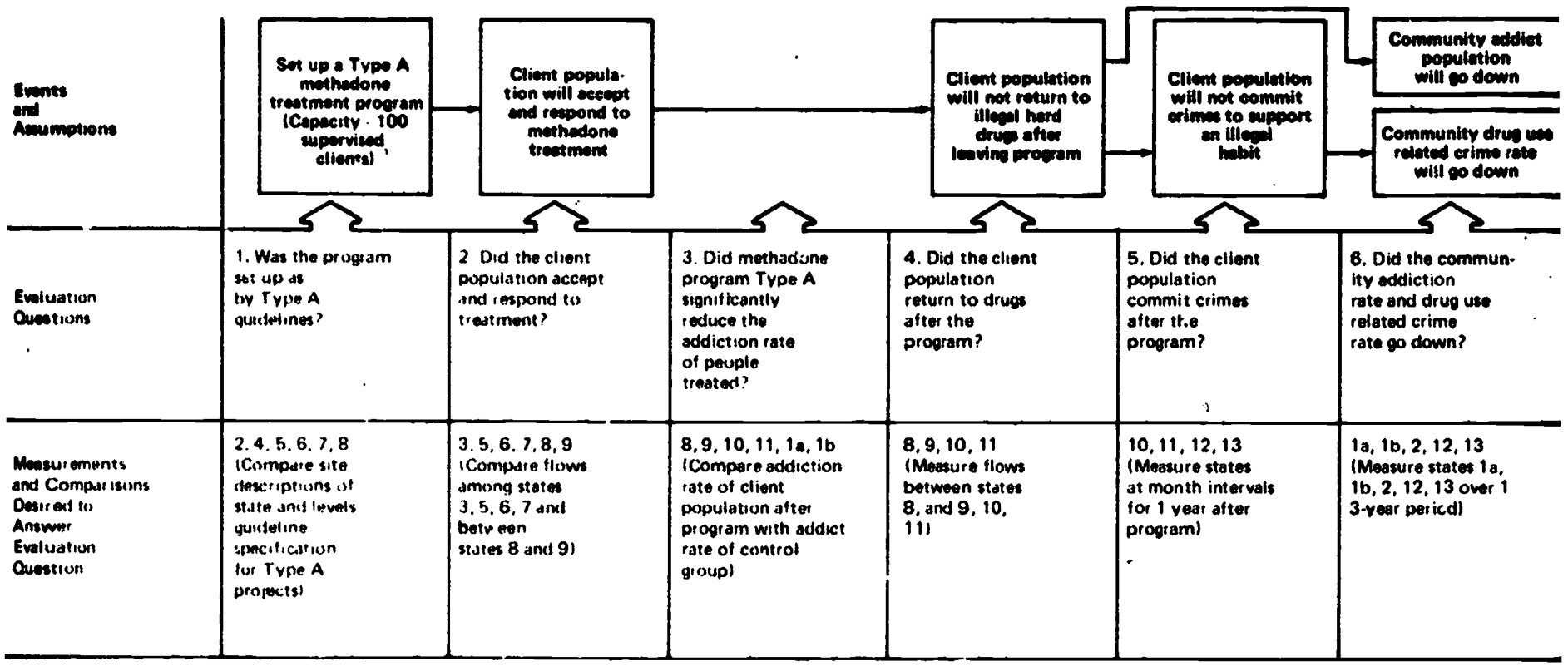


FIGURE 4 Function Model: Methadone Treatment Program (Type A) in City X

FIGURE 4 (Continued)

Measurement Point	Definition of the State of the Affected Populations	Measures of Interest	Measurement and Data Collection System Available
1a, b	Total populations addicted and not addicted to heroin who were never in program	Number of people	No
2	Addicts arrested for any offense	Number	Yes
3	Arrested addicts who are not in program while under jurisdiction of court	Number, rates to and from other states	Yes
4	Arrested addicts referred to program by court	Number	Yes
5	Arrested addicts in orientation and diagnosis component	Number, flows to other states	Yes
6	Clients in supervised methadone treatment	Number, time in component	Yes
7	Clients in counseling component	Number, time in component	Yes
8	Clients in unsupervised methadone treatment component	Number, time in component, rates	Yes
9	Clients in program who return to using hard drugs	Number, time on hard drugs	Yes
10, 11	Ex clients using or not using hard drugs after the program who were never arrested	Number, time in each state	No
12	Ex clients arrested but not using hard drugs	Number	Yes
13	Ex clients arrested and using hard drugs	Number	Yes

FIGURE 5 Logic and Measurement Model for Methadone Treatment Program (Type A)



KEY

Event

Linking Assumption Measurement Point on Figure 2

What Are the Sources of Information for a Program Description?

Since an EA assesses the program management's description of the program, its completion requires management's description plus the descriptions of policymakers, evaluators, and program operators for comparison. Each source may consist of several people, so there may be more than one program description from it.

A program manager at the Federal level is someone with responsibility for those activities established to achieve some national objectives. Whether or not the events comprise a program is arbitrary and a matter of policy. A law, for example, could be broken up into five programs with five managers and five objectives, or set up as one program with one manager and five objectives. One may find a single program with one objective spread out among several departments with several managers who act as one to insure that the national objective is achieved. This diffusion of management responsibility makes EA all the more necessary to bring the pieces together.

Program management, then, is made up of the manager(s) of a program and those people he or she includes in the management decision process. Typically the participants will be an agency director or bureau chief plus the supervisors of line and staff units reporting to them. The management group usually is three to nine people who meet and communicate frequently.

Policymakers are those people with some oversight responsibility for the program, including the power to grant the manager legal authority to administer the program. Even for a major Federal program the policy group usually will not exceed 12 people. Its members are:

- Program Assistant Secretary.
- Under Secretary and or Secretary..
- Office of Management and Budget examiner.
- Government Accounting Office audit division director.
- Staff directors of House and Senate budget committee and oversight committee

For any given program this group may vary. The important point is that in a few weeks one can usually talk to every government official with a significant policy voice in the program.

The policymakers define value for the program—that is, those program outcomes judged to have social benefit. Comparison of their descriptions with those of program management is the basis for the acceptability criterion.

Program operators are those people running specific pieces of the indirect and direct interventions. Their description allows another comparison to estimate validity and plausibility. Program operator descriptions of the program can be pieced together from interviews with the operators responsible for the different activities.

The *evaluators* are the people doing the data collection, analysis, and feedback for the EA—usually a team of two to four people. Their description is based on direct observation. When compared with program management's description, it allows assessment on the validity, plausibility, and cost-feasibility criteria.

Table 5 defines the probable sources of information for each type of model; Table 6 defines the information items carried in each.

Getting Program Management's Description

There are eight steps to complete to obtain a description of the program from the management group. Following is an explanation of each.

Identify Key Actors; Bound the Program

There should be only one program manager and one program involved in an EA. The tasks of defining the key actors and what will be considered as the program are interrelated.

Analysts have had substantial trouble defining the programs of government for two principal reasons. First, the laws enacted by Congress and the subsequent budgetary appropriations are often broad collections of intents, activities, and resources, sometimes covering whole agencies, so that there may be many programs within any single law. Second, the concept of a program is fundamentally a question of definition and perspective as seen by policymakers. The concept implies a set of activities being managed by a single program manager to achieve a common end and, therefore, it has utility only for the purpose of fixing responsibility for achievement.

There are some general criteria which could be applied to the determination:

- The objectives are national, and somehow reflect congressional intent;
- There is a single manager responsible for achievement of the national intent,
- The manager has not delegated all of that responsibility to someone else, and
- There is a defined set of resources and activities presumed as adequate to accomplish the end purposes, and the manager has access to them.

Generally there is a level within a department below which national objectives lose their identity and above which the concept of measurable, achievable ends is lost. That level is where programs are found For example, within DHEW the bureau or institute level is where programs are most likely to be found Above that level, there are agencies or management levels whose leaders are concerned with policy issues and the management of program managers. Below that level, the effort of more than one unit (a division) would be required to achieve national objectives.

The key actors in program management include the individual designated as program manager and his immediate subordinates, e g , a bureau chief and his division directors. In some situations program management might include regional office staff who are directly responsible for administration of some part of the national program

In the best possible situation, the program manager has line authority over the resources needed to achieve the national objectives. In government, however, this authority is often divided. A classic example is the Occupational Safety and Health Administration (OSHA) and its research arm, the National Institute of Occupational Safety and Health (NIOSH). OSHA managers, presumably responsible for administration of the Occupational Safety and Health Act, are part of the Department of Labor while NIOSH personnel report to DHEW A more common example is the situation in which grant administration a major part of a program, has been decentralized to the regional offices, which report administratively to someone other than the program manage In this case, the program

TABLE 5 Typical Sources of Information for Each Program Model Type

Source of Description	Type of Information in the Description		
	Logic Models	Function Models	Measurement Models
Policymakers	Legislation Hearings Committee reports Interviews	Legislation Hearings Committee reports	Interviews
Management	Regulations Guidelines Interviews	Regulations Guidelines Work plans	Guidelines Data system manuals Evaluation designs Interviews
Operator	Interviews	Interviews Work plans	Interviews Files
Evaluator	Observation of program activity and flows	Observation of program activity and flows	Observation of program activity and flows

TABLE 6 Information Items in each EA Model Type

	Logic Models	Function Models	Measurement Models
Definition	The expectations for a program stated as a series of events Events include achievement of the objective and all other events considered necessary to achieve the objective	The flows through a program (such as funds, clients, and information) and the activities that affect them The flows and activities are those necessary to describe both achievements of objectives and operations	A set of measurements and their relationships usually a stock and flow rate characterization The set is a valid representation of the behavior of the function model and describes events in the logic model
Information Items Carried in the Models (Some will be found to be missing during an EA Their absence is to be noted.)	• List events in sequence • Identify measures necessary to define each event (necessary conditions) • Give sufficient definition of each event so that prior events are the cause (The evidence that A causes B is part of the definition of B) • Describe the expected state of the event the conditions that define when the occurrence of the event is satisfactory e.g., excess bed capacity must decrease from 10,000 to under 2,000 • Describe the time frame within which successful occurrence of events should occur	• Name all important flows • Name all states of each flow • Name all important functions • Describe how each function is to affect state of the flow (rate and past states) • Identify flows sufficiently to enable description of operations that contribute to objectives • Identify flows and states sufficiently to enable description of objective achievement or non achievement • Describe how each function is designed to bring about the desired change of state in the flow. Note operating characteristics	• Display of measures (state, stock) and their relationships (flow rate) • Identify the measurement systems providing data • Describe each measurement: - characteristic measured findings - instrument measurement operation estimate of error • Identify where on functional model measurement is taken • Identify which set of measurements are necessary and sufficient to describe each event in logic model • Identify the cost of the measurement system

manager must form agreements with other offices to insure that he gets the needed resources. If he does not have line authority over his resources and cannot negotiate an adequate priority to obtain them from someone else, the program manager must present his problem to upper level managers and policymakers for resolution.

The identification of key actors and definition of the program will bound the work of the EA and make clear by definition whose views will represent program management. Note that we generally exclude from that group other management levels and organizations (the offices of the Assistant Secretary and Secretary, OMB, Congress). Their personnel will be included in the group defined here as policymakers. Their descriptions are also required, but in a separate step.

Collect Descriptive Material

Individuals within the management group must tell the evaluator what specific material should be reviewed. Generally, the necessary material is included in the following:

- legislation
- legislative hearing records (these are often helpful in defining intent)
- program plans
- program budgets
- regulations
- guidelines
- research reports.

It is not unusual to review 30 or 40 individual documents that, in the aggregate, describe the program. Often State and local plans are helpful in putting together a coherent description.

Prepare Rough Models

The condensation and translation of a large body of descriptive material into model format requires a substantial amount of judgment on the part of the evaluator—basically, selecting and sorting for key events and activities. Of necessity, some events will be excluded. There will be several opportunities to verify your selection and arrangement with managers, policymakers, and the like.

Often it is helpful to begin by preparing a program structure, a chart resembling an organization chart but which lists key objectives, subobjectives, and activities needed to achieve the objectives instead of organizational units. Initially, it is easier to prepare a logic model and use it to begin preparation of the function and measurement models. These models need not be presented to anyone at this stage, as they are intended merely to assist the evaluator in constructing his interview guides.

Identify Key Events and Gaps

In final preparation for the interviews with program management, the rough models are reviewed and an attempt is made to identify areas and events which seem to be incompletely described. The focus of the interview should be on these areas.

Interview Key Actors

It will be helpful to prepare a guide before each interview, spelling out the information you want to get from the interview. However, be prepared to listen carefully to everything that is said and take detailed notes. In fact, we have found it helpful to have two people conduct each interview, so that at least one complete set of notes will be produced.

After the interview, the notes should be written up as soon as possible. Of course it is important to try to avoid injecting your view into the notes. Remember that real people with responsibility for the program are your source of descriptions. To insure that the description is accurate, send a copy to the person you interviewed, asking him to confirm it.

The completed interviews should provide enough information to define logic, function, and measurement models. If this is not so, you may have to schedule reinterviews to obtain specific pieces of information.

Prepare Final Models

From all the information at your disposal, prepare logic, function, and measurement models. Remember, though, the models should be treated as a language whose purpose is to communicate. The three models are used because they describe different aspects of the program. What is important is their information content, not their form.

Analyze the Model: How Well Defined Is It?

The criteria and procedure to employ in judging the extent to which the program description is well defined are described in Chapter III.

Verify With Management

Your description (the models) and assessment should now be communicated to management. We recommend a briefing rather than a written report to save time and increase communication. Management should either agree with the findings or suggest specific changes to the description.

Getting Policymakers' Descriptions

Getting a program description from policymakers follows roughly the same series of steps as was outlined for the program management group and is done primarily to determine the differences between the two views. Policymakers, however, are generally a larger and organizationally less related group and include the following members:

- Agency head
- Assistant Secretary
- Secretary
- OMB budget examiner
- Congressional committee members
- GAO representative.

Whenever possible, the actual policymaker should be included but be prepared to settle for proxies, such as an aide to the Secretary or congressional committee staffers (at least one staff member from minority and majority sides for both authorization and appropriations committees from both houses).

The following six steps comprise the activity.

Identify the Key Actors

Get a list from the program manager of the key officials who should be contacted.

Analyze Descriptive Material: Prepare Rough Models

The material collected earlier for the program management description should contain all of the needed documents for the policymaker description. Key documents include legislation and hearing records. Office of the Secretary directives issue analyses, top level plans, and budget.

Again the purpose of the models is to prepare for the interviews. The models may be rough, although often the legislation itself can provide substantial descriptive material of a functional nature. Figure 6 was prepared early in an EA as a functional model, based heavily on examination of the legislation, and is in more detail than most such models derived at this point. It had not yet been validated, but it proved helpful in many subsequent discussions and interviews.

Interview Key Actors

Interviewing policymakers often presents special problems although this is less true for outside consultants than for government personnel. We have always found such interviews to be valuable. However, government personnel have indicated that while consultants may be able to hold such interviews, they do not have the same access to the players. Furthermore, the congressional aides who have indicated that they welcome such opportunities are viewed as relative rarities.

Prepare Final Models

Based on the review of written documentation and the interview notes, the evaluator should prepare final logic, function, and measurement models to the extent possible.

Compare Policymaker and Program Management Models

The criteria and procedures for analyzing the extent to which policymaker and program management descriptions of the program agree are outlined in Chapter III.

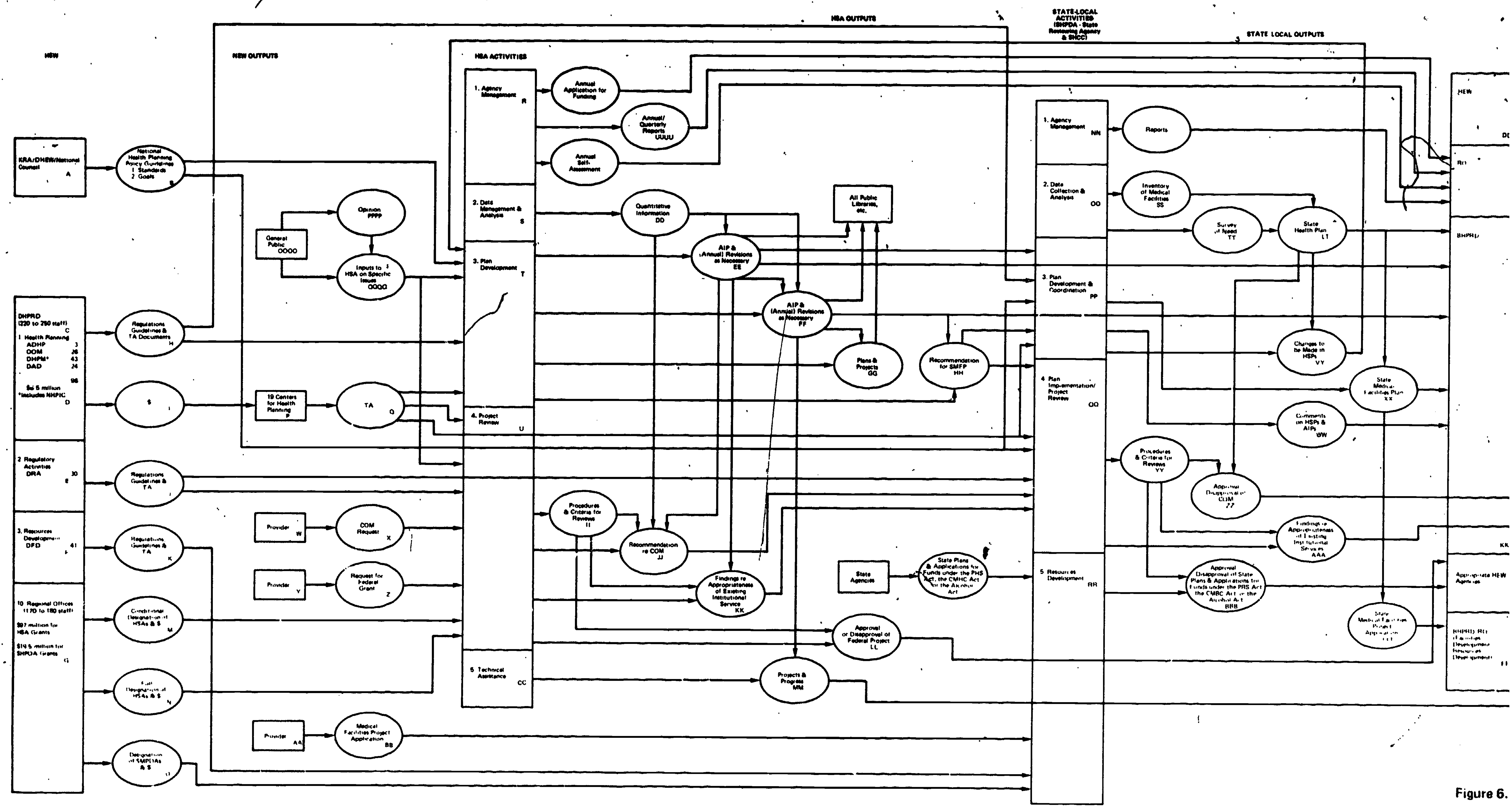
Validate Differences

Whenever differences between the two descriptions seem apparent, they need to be validated. The first point of contact should be with the program manager, who is asked whether he agrees or disagrees with the different descriptions in the policymaker models. Should he disagree, it will be necessary to consult with one or more of the policymakers to be certain that your understanding of their intent is valid. Any policy disagreements discovered by this process will have to be resolved by the program manager with his superiors.

Getting Program Operator's Description

The program operator is typically a grant recipient who provides some service to the public under conditions laid down in the law, regulations, and his individual grant. The program operator might be a local legal aid office director, a health systems agency director, or the director of a State planning agency. In the health planning program (Public Law 93-641), each of the individuals administering State health planning and development agencies and the health systems agencies (HSA's) is a program operator.

Typically, their description of the program will reflect a relatively narrow focus, the perspective of locale. Yet, if the program is to work nationally, the efforts of local program operators must add to the national objective. There is only one effective way to get their description. Go onsite.



HEW
HEW OUTPUTS
HSA ACTIVITIES
HSA OUTPUTS
STATE-LOCAL ACTIVITIES (SHPDA - State Reviewing Agency & SHCC)
STATE LOCAL OUTPUTS
KRA/DHEW/National Council A
National Health Planning Policy Guidelines 1 Standards 2 Goals B
DHPRD (220 to 250 staff) C
1 Health Planning
ADHP 3
OOM 26
DHPM* 43
DAD 24
96
*Includes NHPIC
D
2 Regulatory Activities DRA 30 E
3. Resources Development DFD 41
10 Regional Offices
$97 million for HSA Grants
SHPDA Grants G
Regulations Guidelines & TA Documents H
$
Regulations Guidelines & TA
Regulations Guidelines & TA K
Conditional Designation of HSAs & $ M
Designation of HSAs & $ N
General Public OOOO
Opinion PPPP
Inputs to HSA on Specific Issues QQQQ
19 Centers for Health Planning P
TA Q
Provider W
COM Request X
Provider Y
Request for Federal Grant Z
Provider AA
Medical Facilities Project Application BB
1. Agency Management R
2. Data Management & Analysis S
3. Plan Development T
4. Project Review U
5 Technical Assistance CC
Annual Application for Funding
Annual/ Quarterly Reports UUUU
Annual Self-Assessment
Quantitative Information DD
All Public Libraries, etc.
AIP & (Annual) Revisions as Necessary EE
AIP & (Annual) Revisions as Necessary FF
Plans & Projects GG
Recommendation for SMFP HH
Procedures & Criteria for Reviews II
Recommendation re COM JJ
Findings re Appropriateness of Existing Institutional Service KK
State Agencies
Approval or Disapproval of Federal Project LL
Projects & Progress MM
1. Agency Management NN
2. Data Collection & Analysis OO
3. Plan Development & Coordination PP
4 Plan Implementation/ Project Review QQ
5 Resources Development RR
Reports
Inventory of Medical Facilities SS
Survey of Need TT
State Health Plan
Changes to be Made in HSPs VY
State Medical Facilities Plan XX
Comments on HSPs & AIPs WW
Procedures & Criteria for Reviews YY
Approval Disapproval of COM
Approval Disapproval of State Plans & Applications for Funds under the PRS Act the CMBC Act or the Alcohol Act BBB
HEW
RO
BHPRD
Appropriate HEW Agencies
KK

Figure 6.

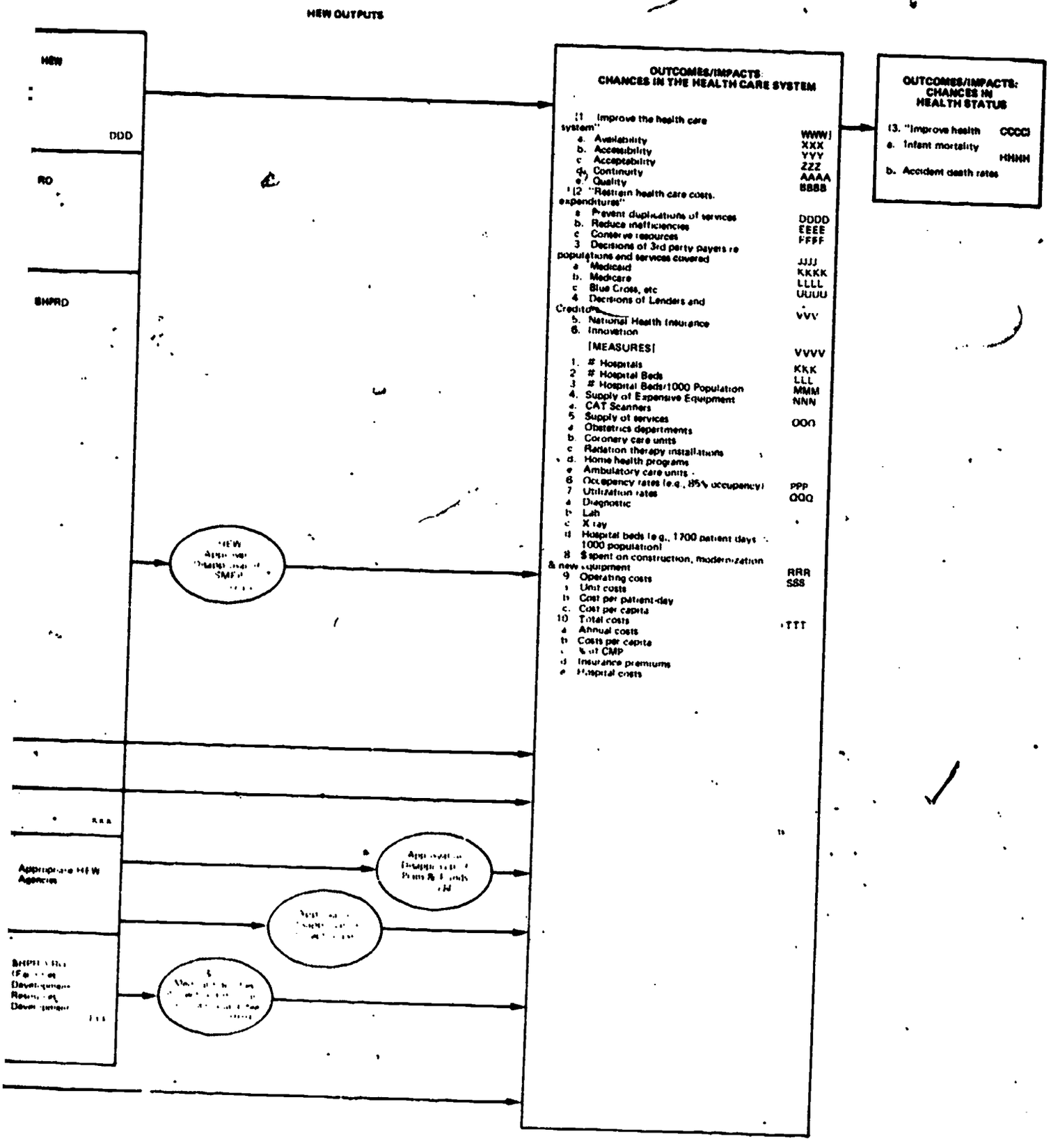


Figure 6. The Health Planning and Resources Development Program (BHPRD and HSA Perspectives): Process Flow Model (with Expected Changes in the Environment)

Selecting Sites

The selection of number and location of sites is important because what is sought is a picture of the program as it really operates. Of necessity, however, one needs to compromise because of the constraints of time and budget. In the health planning program, five sites were visited for 2 days each. In similar work for the Appalachian Regional Commission (ARC), eight projects, six local planning units, and two State program offices were visited to obtain a description of the program from the perspective of local program operators.

The number required depends partially on how much variation is expected in fundamentally different types of activity. For example, in the health planning program, only two types of field organizations were involved—one at the State and one at the local level. For ARC, there were State units, local planning units, and 30 to 40 different types of service providers.

Preparing For the Site Visit

Preparation for the site visit is similar that for interviews with program managers and policymakers In this case, however, the documentation has been predigested. Principally, there will be a grant application, local plans, and the set of models derived earlier, which provide clues to help in focusing the onsite work. The purpose is to determine the extent to which the national program is supported by local programs—in other words, do the two sets of descriptions match?

Review of the available documentation, including models, should be used to determine what to look for onsite Basically, you will be looking for:

- objective statements,
- measurement indicators,
- intervention activities,
- operating assumptions,
- project resources,
- external intervening variables, and
- data sources.

In preparation for the site visit, you should draw up an agenda outlining the objectives of your visit and send it to the people with whom you will meet.

Conducting the Site Visit

Site visits will generally involve two evaluation personnel and last 2 or 3 days (up to a maximum of 5 days for a complicated program). While onsite, the evaluators should brief the director and his staff on the objectives of the visit, which will be pursued by interviewing the director and his staff, observing and recording actual processes, reviewing data, plans, records, and local information systems, and preparing flow charts of the opera)

Before concluding the site visit, it is wise to brief the director on your findings in an attempt to verify your description of his program Try o obtain the following types of information.

- *Objective Statement*--Obtain narrative statements of the objectives of the program as defined by field management personnel. In obtaining these statements, try to get definition of both the immediate output objectives expected as the direct result of intervention activities and the realistic expectations of intent.

- *Measurement Indicators*—Get the management staff to detail their best judgment on what measurement indicators they would accept as plausible success measures of their objectives. The defining of measurement indicators is an iterative analytic exercise involving the evaluator and local program management for the purpose of reaching some tentative agreement on the meaning of the objectives in measurable terms.

- *Intervention Activities*—In obtaining a description of the actual activities being conducted by local personnel, it is important to define the actual intervention activities being conducted onsite, as well as the relationship of activities to each other and to the program objectives. It is expected that local programs may differ substantially in terms of the activities conducted and the manner in which the activities are conducted in relation to other ongoing activities.

- *Operating Assumptions*—While onsite, attempt to obtain the views of management on the chain of assumptions that led management to believe that the intervention activities would lead to achievement of objectives. You will need to add your own assessment and views of these assumptions later, but the management of the program can substantially enhance your understanding of the idiosyncratic conditions that may have led to the adoption of one approach over others that are followed in different sites.

- *Program Resources*—Obtain as accurate an estimate as possible of the resources actually deployed by the program. The accuracy of such information will undoubtedly vary among sites, depending upon the cost records maintained and the extent to which the local program is operated independently of other nonprogram activities. Document the basis used, including records consulted, in determining the allocation of resources to activities.

- *External Intervening Variables*—Attempt to identify any variables outside the control of the program which could plausibly affect the validity of its operating assumptions.

- *Data Sources*—Attempt to define and review the data systems maintained by the local program, obtaining copies of records and reports. The local programs will often prepare internal reports which do not get forwarded to Washington.

- *Project Flow Chart*—Develop rough flow charts that describe the program as a linked set of activities and objectives from the expenditure of funds to the achievement of intended impact. Be careful to avoid introducing or modifying such factors as objectives and causal links, in an attempt to introduce greater logic or tidy up the program design. This task is not intended to describe what should or could exist; rather it must capture what *does* exist. If there are significant differences between what you have been told and what you observe, try to verify your observations while onsite.

Developing a Generalized Program Operator's Description

Using the set of local descriptions developed onsite, devise a general framework for describing the program. This task requires synthesis of the material collected from individual sites into a program

model representative of what was observed in the field. This model is to represent as closely as possible what actually exists, and must emanate entirely from the descriptive material you obtained. It is entirely conceivable that analysis of the individual programs will reveal such disparity among programs that no overall model can capture the observed programs without serious distortion. Some possible outcomes of this task include these observations:

- *Local Programs Are Idiosyncratic*—The field programs surveyed reveal such dissimilarity that an overall model is not feasible;
- *Sets of Programs Emerge*—There exist several models of sufficient similarity that generalized models can be developed; or
- *One Model Predominates*—There exists such overwhelming similarity among local approaches (even if not all programs follow this approach) that one generalized description appears useful.

Compare Program Management's and Program Operator's Descriptions

Criteria and a procedure for examining the extent to which program management's description is supported by, or conflicts with, that obtained from the program operator is provided in Chapter III.

Communicate Findings to Program Management

Using a briefing format, communicate the results of your analysis to program management as soon as possible

The Evaluator's Description

Generally the evaluator will not prepare a separate set of models or descriptions, but rather will make changes to the other models which he is shown. If an evaluator description is prepared formally, it is a composite description, gleaned from the others he has reviewed.

III. THE SECOND STEP: ANALYZE PROGRAM DESCRIPTION

Is Program Management's Description Well Defined?

Rules To Follow

The basic rule is that a description of a program is well defined when it contains all of the elements stated in Table 2, Chapter II. Table 2, as already noted, is a listing of the information needed for a complete program description and criteria to help the evaluator judge the adequacy of the information. The table can be used in the form of a simple checklist (See Table 7), whereby the evaluator can judge, event by event, whether the descriptive material he has at his disposal is adequate; his conclusions can be recorded on the checklist with a simple yes/no.

Based on our experience, the elements in Table 2 constitute the minimum information set needed by an evaluator to complete an EA. Other evaluators may feel the need to add elements which they believe to be necessary. Indeed, in our experience we have found the need for additional information. The desire is to manage a program to achieve specific results.

Procedures To Follow

- Begin the examination of the program description with a checklist, the criteria outlined in Table 2, and the program description.
- Using the model formats, extract needed information elements (using Table 2) from the program description.
- Make judgments on whether or not the program description includes all needed elements. Record conclusions on the checklist.
- Review conclusions with "owner" of program description (e.g., program manager). If new information is offered, alter conclusions.
- Document final agreement between yourself (as evaluator) and owner on final conclusions regarding how well defined the description is.

Using the Models. We have already stressed the value of using models as a kind of shorthand description of what we know about a program. We assume here that the evaluator has already collected as much basic descriptive information as he can, and that he has constructed the three types of descriptive models.

Using the models and a checklist derived from the table of information elements, the evaluator can go through the checklist for each event in the logic model, making yes/no decisions on each element.

TABLE 7 Checklist for Recording if a Program Description is Well Defined

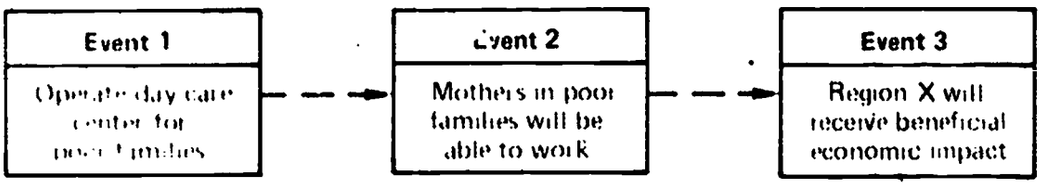
INFORMATION ELEMENTS	Event 1	Event 2	Event 3	Event 4
1. Event sequence				
2. Event descriptions	Yes/No			
3. Measures and comparisons	Yes/No			
4. Expected values for each event	Yes/No			
5. Activities that must occur for event to take place				
6. Information system used to provide evidence				
7. Uses of evaluation information				

For any event in the logic model, it may be necessary to refer to one of the other two models to determine whether the logic model event is well defined. For example, an event in a logic model might be "the cost of hospital care in Region X is reduced." The table of information items asks whether there is a measure and a comparison defining this event. By referring to the measurement model, the evaluator might find a measure, "average per diem cost acceptable by Blue Cross in 1979 compared with the same cost in 1978." For that event in the logic model, then, the evaluator could indicate "yes," the event has a measure and a comparison base. (Note: Subsequent analyses might reveal the measure to be unacceptable for one reason or another. Our concern here, remember, is only with a determination of whether or not the description is "well defined," that is, it contains all of the necessary information elements.)

Similarly, it will be necessary to refer to the functional model to determine whether each event is defined in terms of the activities that must occur for the event to take place. Table 8 summarizes the source by type of model for each of the information elements in Table 2.

Expect to Find Gaps. Going through the checklist for the logic model as a whole and for each event in the model separately provides the evaluator with a picture of the completeness of the description. For any program, it would be extremely unusual to conclude that all information is available especially on the first pass at a management description. The expected condition is that as you move from item 1 (event sequence) to item 7 (uses of evaluation information), more partial information and complete gaps will appear For a new program, for example, there may be no information system and no well defined uses for information. Measures and comparisons also may be only partially thought out, with expected values defined by "increase," "decrease," etc., rather than specific quantitative values. The functional logic may be only partially thought out. (What events affect per diem hospital costs and how are they calculated?)

The Significance of Gaps. Several points should be made here. First, identified gaps at this stage may prove to be unimportant on further analysis. As an example, consider a simple case of a 3-event model



Via the checklist, we conclude that events 1 and 2 are well defined, but event 3 is not because there is no measure or measurement system. Suppose, though, that we check with the policymakers and they declare that they have no concern about event 3. If we can demonstrate success at the event 2 level, that is adequate. The information gap disappears, and the event should be eliminated from the scheme

Gaps can also be filled in by subsequent work. For example, a gap in defined measures for some event in the management description might be filled in after a visit to program operators in the field reveals that they know how to (and do) measure the event. The point is that before you plug every gap, check first to make certain that (a) it is a gap, and (b) it is important.

Second, there are two apparently different types of gaps: (1) those in the logic or function of the program design itself; and (2) those in the information part of the program (the absence of a measurement system). The evaluator might consider the second type his province; the first type someone else's On closer examination, however, the distinction gets fuzzy. The evaluation design is an integral part of the program design, and the evaluator becomes part of the program design process while he is completing an EA In some areas, to be sure, the evaluator lacks the technological expertise to fill gaps, but in others, he will be the expert. In either case, he should be prepared to participate in the process of defining alternative program design options when necessary.

TABLE 8 Elements of a Program Description that can be Extracted and Displayed by Different Models

Information Element	Logic	Measurement	Function
Event sequence	X		
Event descriptions	X	X	X
Measures and comparisons	X	X	X
Expected values	X		
Activities that must occur			X
Information systems	X	X	
Intended uses of information	X		

Example

The following is an example of how we examined the extent to which the manager's description was "well defined" in one program. One part of the Bureau's health planning program is its technical assistance strategy, illustrated in Figure 7 as a logic model. The boxes represent events that were expected to occur, the arrows indicate causation. The strategy assumed that the Bureau would develop products (e.g., planning methods) which the health system agencies would use to achieve changes in the health care system.

Figure 7 has been simplified for purposes of this example. It represents the actual logic model constructed to represent the Bureau's expectations, as confirmed by management. The assumptions represented by the model would be tested by an evaluator. Do the events occur? Are expectations met?

It is important to note that this model is not a set of research hypotheses dreamed up by an evaluator but rather a policy statement—the logic used to justify the Bureau's technical assistance program to Congress, OMB, and DHEW.

Regarding the question of how well defined is this logic model, a simple yes/no checklist presents the conclusions reached after examining the program description (Table 9). The "yes" boxes signify that the event description satisfied the required information elements; the "no" boxes reveal partial or complete gaps

Table 10 summarizes the information available on the technical assistance (TA) program. The analysis revealed that program design had never included the last two events, except to name them. Missing from the design were processes whereby agency use could be defined and therefore be known. Also missing was a process which could be used to track changes in the agency caused by the program products and link those changes to health care system changes.

This program was judged to be unevaluable on the last two key events. Note that this finding would not prevent the program from being implemented, but rather that it indicated there was no way to know whether or not the TA was useful. Instead of simply communicating this finding to management, the evaluators worked with Bureau staff to define an alternative design which satisfied the criteria of a well defined program (see Chapter IV).

Note that the gaps in the last two events were made more real by the agreement between policymakers and program managers on the need to measure success at the level of events 3 and 4 in Figure 7. Had policymakers and program management agreed that success was defined by event 2, the gaps would have disappeared

Is Management's Description Acceptable to Policymakers?

The purpose of this assessment is to determine whether there is agreement between policymakers and program management on the substance of the program. One should not expect the same level of detail in both, but rather that they agree on the main intent and strategy being followed.

Rule To Follow

The rule to follow is that if events in the policymakers' program description also appear in program management's description, you can assume they agree.

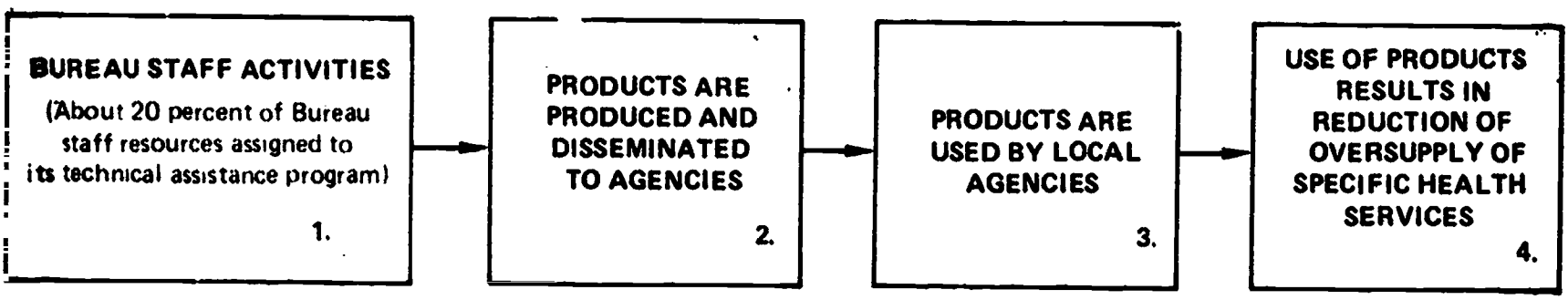


FIGURE 7 Expectations for Bureau's Technical Assistance Program

TABLE 9 Illustrative Checklist: How Well Defined is Program Description?

INFORMATION ELEMENTS Event Sequence Event sequence exists. Sequence starts with event or management control and ends with important objective.	BUREAU STAFF ACTIVITIES (About 20 percent of Bureau staff resources assigned to its technical assistance program) →	PRODUCTS ARE PRODUCED AND DISSEMINATED TO AGENCIES →	PRODUCTS ARE USED BY LOCAL AGENCIES →	USE OF PRODUCTS RESULTS IN REDUCTION OF OVERSUPPLY OF SPECIFIC HEALTH SERVICES
Measures and comparisons	Yes	Yes	No	No
Expected value for each event	Yes	Yes	No	No
Activities that must occur for event to take place	Yes	Yes	No	No
Description of operation used to take measures	Yes	Yes	No	No
Who will use evidence of event occurrence?	Yes	Yes	No	No
Information system used to provide evidence	Yes	Yes	No	No
Uses of evaluation information	Yes	Yes	No	No

TABLE 10 Using the Checklist to Determine if the Technical Assistance (TA) Program is Well Defined

INFORMATION ELEMENTS 1. Event Sequence and 2. Description	BUREAU STAFF ACTIVITIES (About 20 percent of Bureau staff resources assigned to its technical assistance program)	PRODUCTS ARE PRODUCED AND DISSEMINATED TO AGENCIES	PRODUCTS ARE USED BY LOCAL AGENCIES	USE OF PRODUCTS RESULTS IN REDUCTION OF OVERSUPPLY OF SPECIFIC HEALTH SERVICES
3. Measures and comparisons	Staff time and contract by activity: plan vs. action	Products produced vs. products disseminated	No measure	No measure
4. Expected value for each event	Budget: 356 mm $1,973,000 Divided by TA to agencies	All products approved and disseminated	No measure	No measure
5. Activities that must occur for event to take place	Activities defined in (Major Initiative Tracking System) MITS	Activities defined in MITS	None defined	None defined
6. Information system used to provide evidence	MITS	MITS	None defined	None defined
7. Uses of evaluation information	Managing staff resource/ activity plan	Managing staff resources and product activity	None defined	None defined

The two descriptions will have been obtained independently, and therefore, differences between the two may or may not reflect disagreement. If the difference is that events appearing on the policymakers' description do not appear on program management's, further interviews should be held with program management to determine if the events were merely overlooked. On the other hand, if events on the program management's description do not appear on that of the pc "cymakers, the decision on whether to reinterview will depend on what is reflected by the omissions—differences in level of detail or in apparent intent or strategy. A simple checklist can be used for this task.

Procedure

The basic procedure to follow is to (a) compare the logic models of the policymakers and program management, (b) conduct any reinterviews that seem necessary to validate differences, and (c) hold informal briefings for the purpose of highlighting differences and determining whether they reflect disagreement in intent or strategy.

The evaluator is here trying to obtain agreement on a common description so that he can proceed with the EA. Disagreements may require the collection of further information until agreement on one model is possible.

Example

Figure 8 is an example of a program description (presented as a logic model) based on interviews with a policymaking group. Comparison of this logic model with management's description led us to conclude that there was agreement:

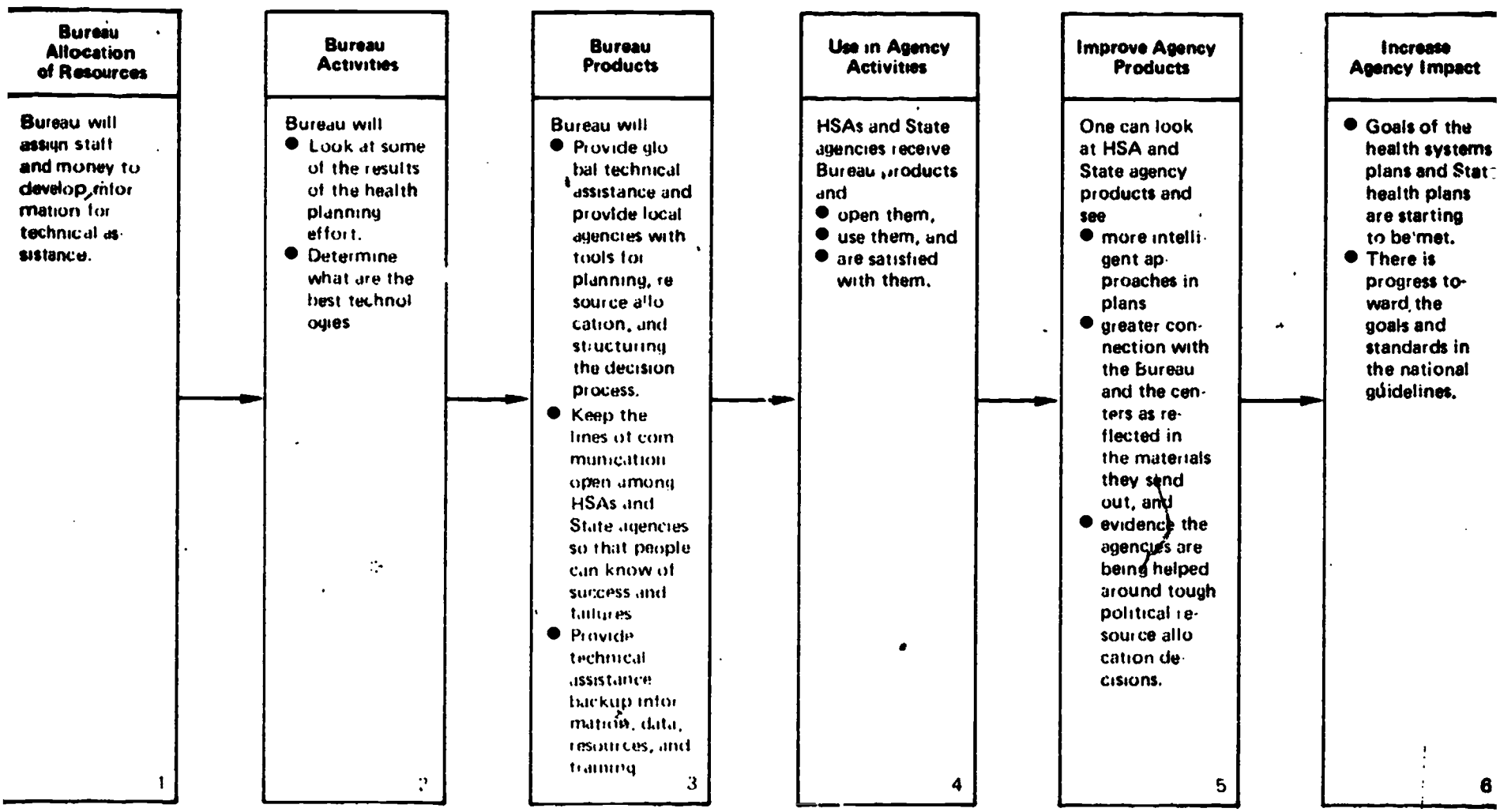
> Those above the Bureau agree with the Bureau in their ·pectations for Program III. Using the results of our interviews with Department of Health, Education, and Welfare (DHEW), Office of Management and Budget (OMB), and congressional staffs, we were able to construct Figure 8, a model of the pr·gram from the perspective of those above the Bureau. This model is similar to the Bureau's view of Program III.
>
> Both the Bureau and those above the Bureau expect Program III to affect significantly the operations and capabilities of health systems agencies (HSA's) and State agencies.

Figure 9 illustrates the logic of this program as now reflected in program management and policymakers' descriptions, and our own assessment of how well defined was the logic. Note the confirmed gaps in the logic—Events 3 and 4 have disappeared. Had policymakers and program managers not both agreed on the need for these latter events as evidence of success, the gaps would have been unimportant. As it was, they had to be filled.

It would be helpful at this point to be able to cite some instances of disagreement between policymakers and program management. Unfortunately, in the few cases in which EA has been applied fully, few disagreements of this type have occurred. One program in which a difference emerged was in the Appalachian Regional Commission's (ARC) health and child development program

A substantial part of the rhetoric defining the Commission's program, especially in the area of health and child care, suggests that the Commission is intended to fund new ideas, not merely to supplement DHEW funding. Interviews with congressional aides revealed that congressmen took this view of the Commission's program seriously, and did not believe that the Commission had made a serious attempt to do anything very experimental.

FIGURE 8 Developing and Disseminating Discretionary Technical Assistance Products, Logic Model of Policymakers' Description*



* Based on interviews with staff members of DHEW, OMB, and congressional legislative and appropriations committees

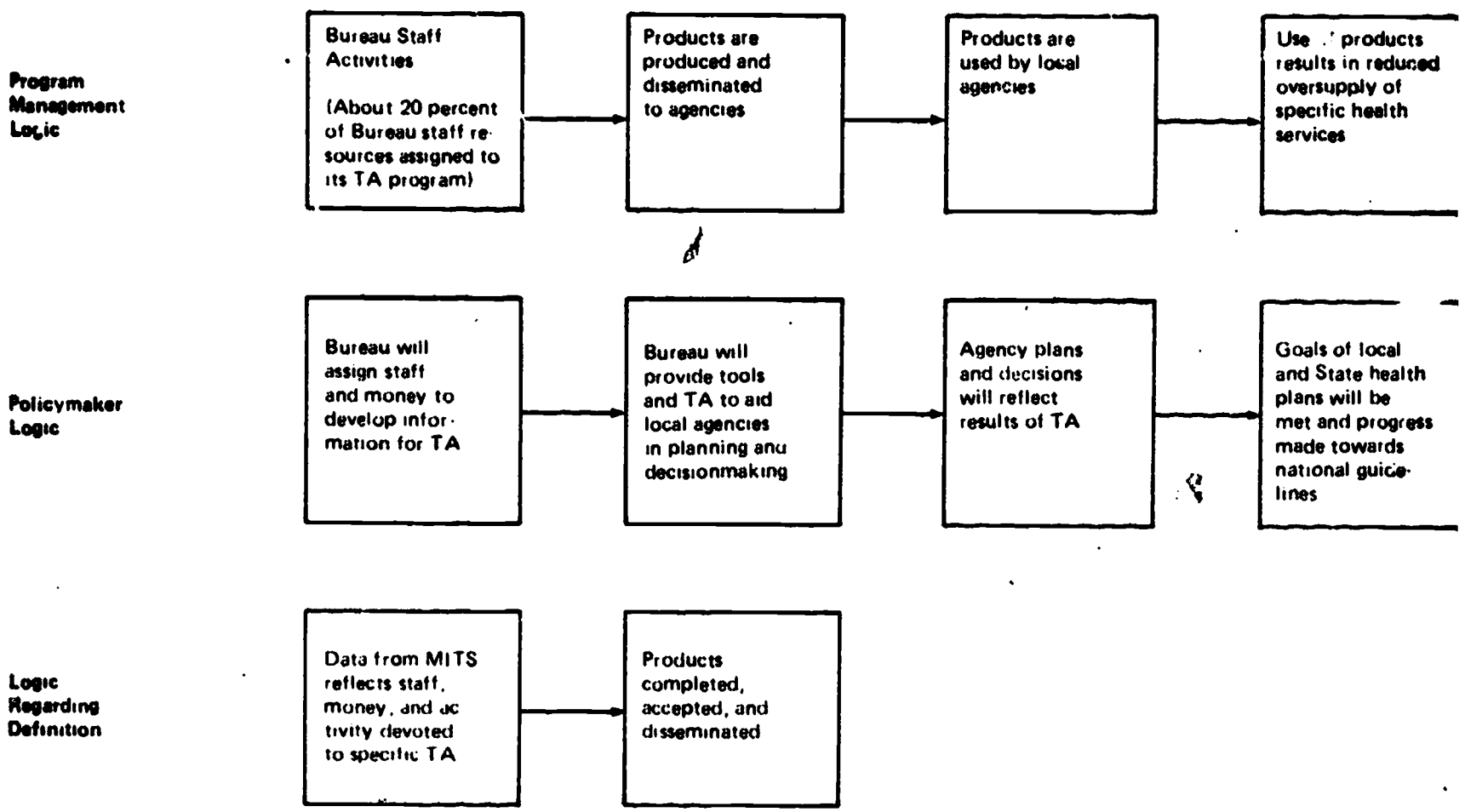


FIGURE 9 Comparison of Program Management and Policymaker Logic with Assessment of How Well Defined was the Program Description

Figure 10 (Schmidt, et al. 1976) illustrates the logic models defining the innovation objective. Analysis of the model indicated that the objective itself had never been defined in measurable terms, and that the Commission's statistical reporting system by design did not segregate innovative from conventional projects. The Commission had never developed any processes for implementing the idea.

The situation as revealed by the EA, was that Congress had expectations not shared by the program management group within the Commission. Note that the EA did not conclude that the Commission failed to sponsor any innovative ideas: in fact, it did. The conclusion was only that there was no formal process whereby innovations were separately identified, no overt incentives to encourage the development of innovations, and no information processes to study and disseminate information about innovations.

A tentative alternative was developed to add an information process. The alternative was subsequently adopted and subjected to further analysis to define the objective and an operational approach, thus bringing the Commission's program more nearly into line with the policymakers' (Congress) views.

Is Management's Description a Sound Representation of the Program in the Field?

The spatial and organizational distance that separates Washington from real program activities in the field creates a special problem for managers and evaluators. It would not be unusual to find arguments raging in Washington about objectives, strategies, and the fine points of evaluation research that simply disappear when program operators meet and service the public. Before attempting to answer the question, "Does it work?" it is wise to ask, "What is it?" Does the program debated in Washington exist in that form in the field?

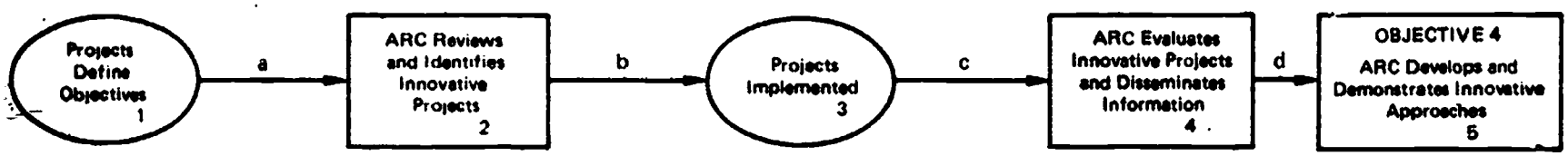
Rules To Follow

Table 2 identified the sets of information that constitute a program description for management and evaluation purposes. Two sets describe the program activities which must exist onsite to produce results: measures and comparisons defining each event; and activities that must exist for events to take place.

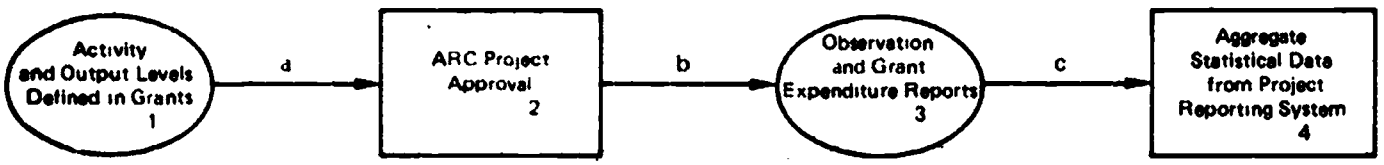
This part of the EA estimates whether or not management's description of these events accurately represents the program as it operates. The rule to follow in making this determination is that management's description of the program is valid when the field activities observed and described by the evaluator conform with management's description of them, and management's measures and comparisons are sufficient to describe accurately the activities there.

The rule is applied by comparing management's description with one or more independent descriptions of the activities. Since the operator is closest to the program and the evaluator's description represents an impartial, direct observation, these two descriptions are likely to be correct. When disagreements arise, additional onsite observation can determine which of the three descriptions is valid.

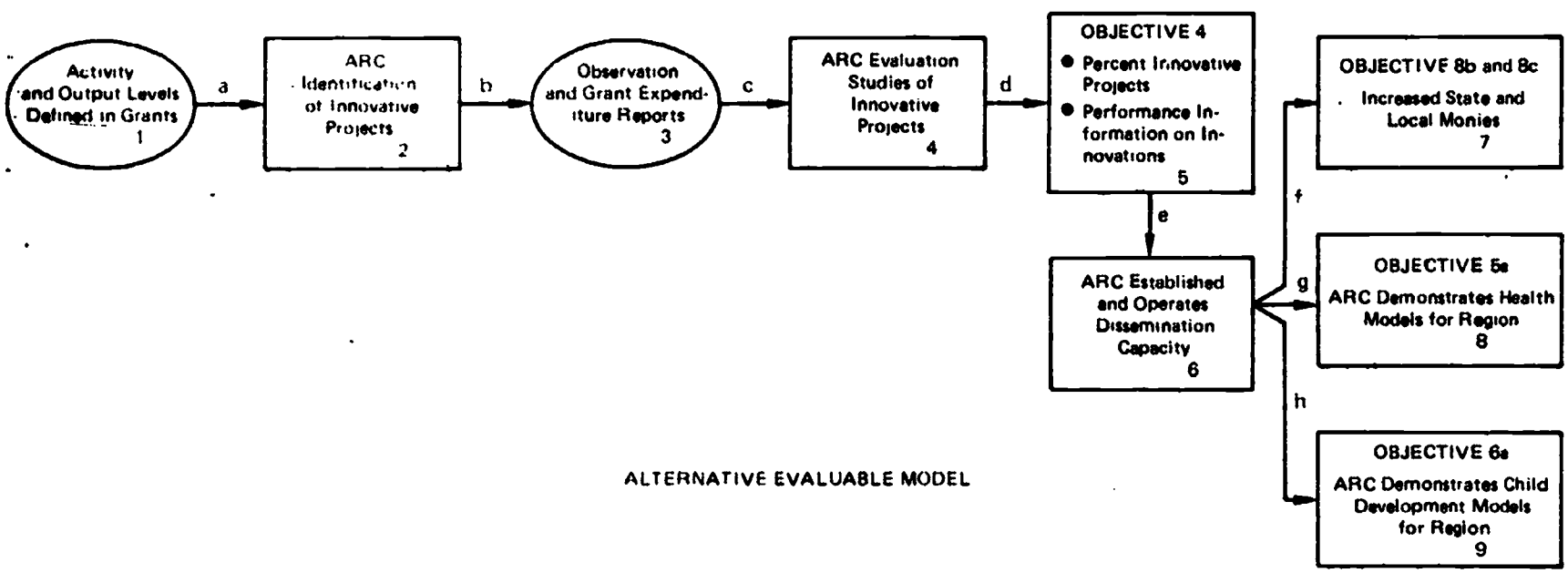
Table 11 is a checklist to record the conclusions reached in applying the rule. The evaluator should be prepared to support each conclusion by presenting material from the appropriate program description.



RHETORICAL PROGRAM MODEL



EVALUABLE PROGRAM MODEL



ALTERNATIVE EVALUABLE MODEL

FIGURE 10 Program Model for Demonstrating Innovation in Appalachian Regional Commission (ARC) Project

TABLE 11 Checklist for Comparing Program Management's Description with the Program in the Field

Event Sequence →	Event 1 →	Event 2 →	Event 3
1. Program management has identified activities that must be conducted onsite	Yes/No		
2. Program management has identified measures and comparisons	Yes/No		
3. Management's activity descriptions are valid representations of program activity onsite	Yes/No		
4. Management measures and comparisons are valid representations of activity onsite	Yes/No		

Procedure To Follow

- Take management's description of program activity and the measures and comparisons associated with each event identified. (Measurement and function models can be used to facilitate extraction of these elements from the description.)

- Compare management's description of activity with what operators say exists onsite and with what the evaluator has observed onsite.

- Make a judgment as to whether management's description is a valid one. Record the conclusion in Table 11.

- Compare management's description of measures and comparisons required to describe an event with what the operator and evaluator believe is required to capture the behavior of actual activity onsite.

- Make a judgment as to whether management's measures and comparisons are sufficient. Record the conclusions in Table 11.

- Review conclusions with program management and note areas where management agrees, disagrees, or requires further evidence. Continue providing evidence and discussions until there is agreement between management and evaluators on conclusions.

- Document final agreement.

The program operator's (or evaluator's) function model represents the activities observed by the evaluator to exist onsite which should be compatible with that of management. Differences should be verified by the evaluator to insure that the field models are correct.

The program operator's (or evaluator's) measurement model represents those sets of measures which the evaluator believes would represent the activities and outcomes of the observed program and those sufficient to define the events of interest to management. Again the operator's measurement model should be compatible with that of management, and verification may be necessary should important differences be observed. Conclusions regarding measurement potential (those measures common to both) and measurement problems (differences observed) can then be reported

This particular comparison can be used to avoid an error frequently made by evaluators—measuring something which does not exist. It is an area in which cleverness and a compulsion to produce elegant analyses can lead evaluators to define their own measures and measurement systems which may not represent the real activities and intent of a program.

Example

Evaluation studies completed for the Atlanta public school system provide an example of differences between a program operation as perceived by management and as it really operated. In this case neither spatial nor organizational separation was as severe as is encountered in most Federal programs yet differences in perception were still significant. Note that the Atlanta example is an instance in which EA criteria were not used until near the end of the project, and then only to find out why the system did not work as intended Consequently, we failed.

Figure 11 illustrates a simplified logic model of the school system. It shows the school administration making decisions regarding subjects such as assignments of teachers and other resource personnel, training, and various sanctions and rewards. These processes are expected to affect teacher and overall school performance, reflected by student performance measured through standardized tests.

A relatively simple signaling system was devised, whereby test results from classrooms were plotted and converted statistically into above and below average performance for students from relatively comparable economic backgrounds. These results were translated into signals—symbols representing student performance in various schools and classrooms. It was assumed that such information would be used to target the various resources.

The evaluators did not develop function models of the actual processes and decisions used to make the assignments, and it was discovered afterward that performance had not been and still was not used as the basis for making these decisions. Table 12 illustrates a checklist that might have been prepared for Altanta had we conducted an EA. Note the areas in which management's description proved to be incorrect or incomplete. Had we known of that problem, we would have been able to stop at that point until we reached agreement with management on the need to introduce new activities and measures

To use such a performance signaling system, the school administrators would have to develop and adopt a different method of operation. First, the signals would have to be used as triggers releasing personnel skilled in classroom performance assessment. Next, these personnel would have to compare teaching styles and skills to determine what factors seemed to contribute to above or below average performance in specific classrooms. Next, correctives would have to be developed and applied Finally, performance changes would have to be measured. These functions did not exist, and consequently, the evaluation system failed because the evaluators did not verify the validity of management s description of the system actually used to correct performance problems.

It should, however, be noted that a different interpretation is possible here. Everyone knew that resource teachers were not assigned according to performance criteria, but it was assumed that the criteria would change with availability of the signaling system. The evaluation demonstrated that the criteria did not change, and therefore, it was successful in pointing out this fact. The authors believe this outcome could have been predicted and avoided by designing new activities and criteria.

Are Management's Expectations Plausible?

However well defined is the description of the program, judgments about plausibility remain. Many have argued that plausibility judgments should not and cannot be made until after a carefully designed evaluation study has been completed, but we feel they should be made as early as possible and preferably in advance of an evaluation study. We maintain that arguments against such analyses are akin to taking a stand against the need or possibility of completing technical feasibility studies in advance of building a bridge

Rules To Follow

Two of the elements in Table 2 describe how the program is expected to perform: event sequence and expectations The plausibility analysis in EA asks what is the likelihood these expectations will be realized with the program in place or being set up? If the likelihood is high, the expectations are labeled plausible, if the likelihood is low, they are implausible.

The rule to follow in making this determination is that management's expectations are plausible when there is some evidence that the program activity onsite will achieve the results expected, and

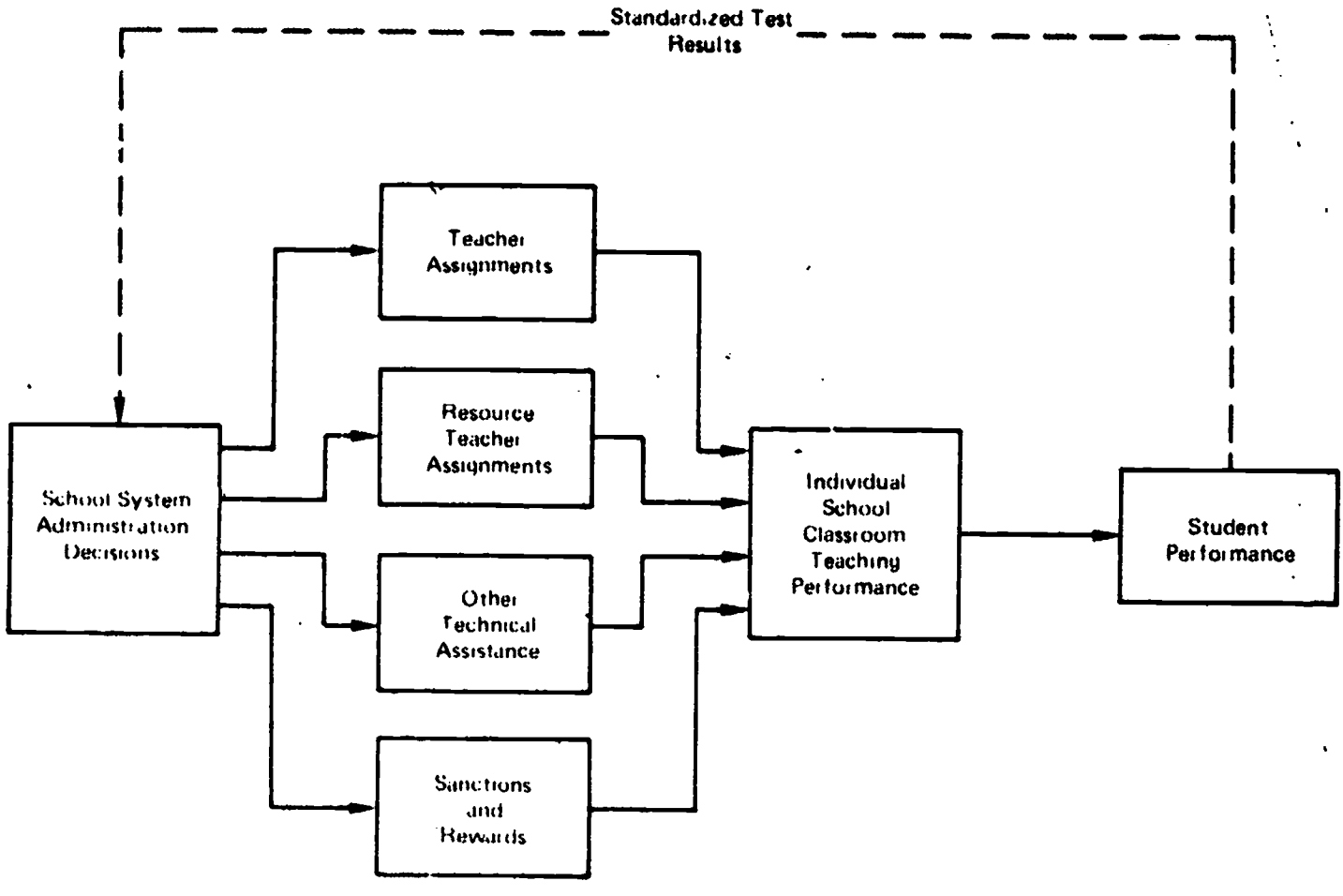


FIGURE 11 Logic Model for the Administration of the Atlanta School System

TABLE 12 Checklist for Atlanta System

Event Sequence	Teacher Would Teach →	Signals Would Identify "Best" "Worst" →	Investigation Would Uncover and Fix Things →	Overall School Performance Would Improve Investigation and Intervention
1. Program management has identified activities that must exist onsite	Yes	Yes	Yes	Yes
2. Program management has identified measures and comparisons	Yes	Yes	Yes	Yes
3. Management's activity descriptions are valid representations of program activity onsite	Yes	Yes	No	No
4. Management measures and comparisons are valid representations of activity onsite	Yes	Yes	No	No

there is no evidence to the contrary The types of evidence that can be used are research or evaluations showing what results these activites produced in the past, the opinions of experts and operators. and the judgment of whether the program activity is large enough to create the expected result

Table 13 is a checklist to use in applying the rule. The basic rule requires observation and judgment The key is to look for evidence that the program is plausible, with the absence of convincing evidence making an argument for implausibility. Remember that "implausible" is not a judgment that something *will not* work; it is merely an argument that there exists either no or inadequate evidence that it *will*.

Procedure To Follow

- Management s event descriptions and expected values for events are associated with performance (output products, expected results).
- Compare management's expected values with those obtained from the operator or evaluator description.
- Make a judgment as to the relative plausibility of management's description
- Record the conclusions reached on Table 13. Record evidence used to support conclusion.
- Review conclusions with program management and note areas in which disagreement exists. If needed, obtain additional evidence from the field and continue to work with management until agreement is reached on a plausible description.

Assuming the two sets of event descriptions correspond, the basic test is performed by comparing expected values for each of the events. Note that the program operator's description to which we refer should have been converted to an aggregated description, especially regarding expected values The issue is that assuming all programs operated in the field perform at their expected values, will their aggregated performance meet or exceed the expected values found in the management description?

Conclusions are obviously easier to reach in cases where necessary activities are not in place at all Although such situations have in fact been observed, the more likely situation is one in which sufficiency is the issue or in which the technology is questionable.

Wherever threats to achievement appear, the evaluator should reconfer with program operators to be certain that the apparent threat is real. The evaluator should note for each event whether it is plausible or implausible, and what is the evidence for making the determination.

Obviously analysis of plausibility is likely to be different for new and ongoing programs. The most obvious difference lies in the existence and relative experience of program operators. Brand new programs will have no current program operators, so a vital comparison is missing. In such cases, the evaluator will have to rely on review of the literature supporting the program, on extrapolation of management s expected values from single sites (in a program where many sites will contribute), and on the analysis of resources likely to be available for the program While the available information will be less and the conclusions necessarily more speculative, the analysis is no less useful in such situations

Example

The findings that might emerge from an analysis of program design plausibility relate to the general notion of the necessary and sufficient set of conditions that must exist for a program to succeed.

TABLE 13 Checklist for Plausibility Analysis

Event Sequence	Event 1	Event 2	Event 3	Event 4
Expectation				
1. There is evidence of plausibility (yes/no)				
2. There is evidence it is not plausible (yes/no)				
Plausibility conclusion (yes/no/questionable)				

Necessity argues that some definable set of conditions, both internal and external to the program, must exist. Sufficiency argues that these conditions must be present in the right quantity. Conditions are resources (money, people, facilities, equipment), activities, and other factors affecting performance.

One example of an EA leading to doubts about plausibility concerns the Bureau's technical assistance program—a subprogram of the overall health planning program (Public Law 93-641). The conclusion emerging from the EA was that this program was unlikely to result in the occurrence of certain events (specifically, events 3, 4, 5, 6 in Figure 8). Here the analysis produced certain conclusions about the extent to which the necessary set of conditions existed. The conclusions rested on several observations:

- Interviews with Bureau staff revealed that they were unsatisfied with their own rate of producing information and the information products themselves;
- Interviews with HSA staff—the main recipients of the Bureau's products—revealed that the products were not highly thought of; and
- Examination of the process being used by the Bureau revealed no activities designed to promote use or to assess the relative effectiveness of the TA effort.

Each of the findings was documented in detail and presented to management, which was asked to decide whether they agreed with the findings, and to determine whether the available evidence was sufficient to warrant action. Further, some alternatives were offered (see discussion in Chapter IV) to stimulate their thinking.

A second example is available within the same program. As a result of the EA, Bureau management agreed that an interim objective for the program was HSA development of evaluable programs to produce national guidelines. It was assumed that development and implementation of evaluable programs was one of the necessary conditions to achieve success, but examination of the plans developed revealed some disturbing findings.

For one, HSA's were adopting objectives, and presumably expending some of their resources, in hundreds of areas. Two, although the plausibility of individual program plans could not be assessed conclusively, the information in the HSA documents strongly suggested that many of the objectives were not supported by real activity. Finally, the tentative conclusions about plausibility conformed with earlier interviews with HSA directors who stated that they knew many of their objectives were implausible, but that they were in the plans to conform with DHEW planning guidance.

On the basis of very limited evidence, we projected that the Bureau would in several years be unsatisfied with the program's achievement. We suggested that there would be considerable anecdotal evidence of small successes at the local level, but that there would be limited evidence of success in areas of national interest. We speculated that the lack of limited and clearly defined priority areas and planning guidance which stressed the need to complete "comprehensive" planning had led HSA's to a position in which their efforts were spread thinly over too many areas.

We suggested the need to focus HSA's more narrowly on areas of national interest, as well as to focus the Bureau's information collection process to access information in stages, with each stage being more narrowly tuned on selected areas than the preceding one. The first stage, for example, might use a telephone survey of a sample of HSA's to determine areas in which HSA program activity promised to be successful. Successive stages might then involve collecting more detailed information about the potentially successful local efforts. Even the initial stage, however, could be made more specific by selecting priority areas of national importance as defined by Bureau policymakers.

The strategy was intended to produce information about those parts of the program which seemed on the basis of objective evidence to be plausible, thus reinforcing those parts. The important point here, though, is that even limited evidence of plausibility can cause management to redirect its activities. In this case, many people within Bureau management were already predisposed to such a finding. In other cases, more conclusive evidence might have been demanded. The key is to feed back such evidence as is available, noting its limitations. As an evaluator, you cannot assume that management will demand conclusive evidence before acting.

Is Management's Evidence Reliable?

In preceding steps, we have determined that management's measures are valid, that is, they measure the right variables. We must now be concerned with the extent to which the evidence to be used is reliable. If it is, the instrument and measurement process will produce the correct values for any subsequent management actions to produce the desired result. Repeated measurements by independent observers which produce the same findings increase management's confidence in the reliability of the measurement system.

Rules To Follow

Management s evidence is likely to be reliable when the following conditions exist:

- There is an independent measurement taken for comparison, prior to action based on information;
- The information is used in such a way that several observers can comment on reliability; and
- There is a capability to get additional information to resolve rel lity issues.

Procedure To Follow

The basic procedure to follow is:

- Examine management's description of the measurement processes to be used for each major event.
- Ascertain whether there exists a process for obtaining repeat observations of events prior to any actions or decisions by management.
- Interview management to determine how many repeated observations are necessary to trigger management action.
- Record judgment as to acceptability of measurement process relative to management's intended use of data.
- Communicate judgment to management.Propose alternative reliability measurement processes if required.

Example

Legal Services Project Evaluation. The Office of Economic Opportunity (OEO) Office of Legal Services implemented a project evaluation system involving review of the quality of legal work performed by its grantees. The procedure involved use of an instrument for recording observations, a team of trained legal observers familiar with the program, and a sample of case files from attorneys in the grantee's office.

Cases were selected randomly from the attorney's open and closed case files. Using an interview checklist, attorney-raters interviewed the attorney about each selected case, and numerical quality ratings were assigned. Two raters were used to produce a consensus rating for each review.

Then, before any management action was taken, successive, independent assessments were conducted by regional office staff and a second site visit team. Further, reactions to the findings were obtained from the grantee.

Based on the findings, one of two management actions was to be taken in each case. For high quality grantees, grants were to be expanded to provide more service, and grantee staff were to be used as sources of technical assistance. For low quality grantees, grants were to be terminated. In such cases, assessments were repeated over a period of more than 1 year, and *only* cases in which grantees showed no improvement were grants canceled or not renewed.

In this example, the instruments—rating sheets and interview guides—were incidental, their reliability unknown. The key to the evaluation was attorney-rater judgment, and the process of consensus-seeking through repeated, independent reviews was crucial. Also, an important factor was that the evaluation process only had to be sufficiently accurate to detect very good and very bad quality. Had it been used to rank order performance of all grantees, its reliability might have been questioned. But since most grantees performed in the midrange, the risk of making a wrong decision here was quite low, given repeated observations and adequate time to correct deficiencies.

Solid Waste Collection. Another measurement system involved assessment of the overall effectiveness of municipal solid waste collection programs. As one part of a larger assessment process, a process and instruments were developed to rate the cleanliness of city streets and alleys. The applied measure was the percent of streets and alleys rated satisfactory on a visual rating scale (Hatry, 1977).

Here ratings are based on visual ratings by a specially trained observer. The cleanliness and appearance of a street or alley are graded in accordance with a set of photographs and written descriptions that cover a range of litter conditions generally found throughout the community. Trained observers driving along the streets (or through the alleys) assign to each block face a numerical rating that corresponds most closely to a grade described in the photographs and written descriptions. The observer does not have to leave the car to make the ratings.

These ratings constitute a readily unders[illegible]neasure, especially when results are presented with photographs on which the rating scale is based. Under proper supervision, the trained observer program provides a reliable way to measure aesthetic impacts on the community and changes in cleanliness over time. Photographic rating scales for cleanliness have been used by the governments of Washington, D.C., New York City, Savannah, Nashville, and St. Petersburg. (Further information on the New York and Savannah efforts is available in reports from both these cities.) The information thus gathered can be tabulated to indicate cleanliness by neighborhood, sanitation service area, or for the entire city. The visual rating procedure, despite its attractive features, can quickly degenerate into an unreliable procedure unless proper care—particularly in supervising inspectors—is taken. Over time, inspectors may forget the definitions and become careless in their application or even in making ratings. In one city, the ratings of one inspector gradually retreated into so narrow a spectrum that he did not use the extreme ratings regardless of the circumstances. (This

scale compression is relatively common when inspectors do not adhere to photographic or written standards.) A second city was unable to use its ratings after it discovered that independent but virtually simultaneous inspections of the same scenes had resulted in wide rating disparities that had not been promptly corrected by retraining the inspectors.

Experiences to date indicate that several actions are needed, particularly after implementation of the rating system, to insure that the data are reasonably reliable. First, the development of the visual, photographic rating scale should be undertaken systematically to assure that the rating scale can reliably distinguish the different levels of cleanliness. Second, adequate training of observers in the use of the rating procedure, both for new observers and as periodic retraining for experienced observers, should be provided to prevent deterioration in rating skill.

Finally, regular checking of sample ratings should be done to determine if observers maintain sufficient accuracy in their ratings. Perhaps 10 to 15 percent of each inspector's ratings would be replicated by the supervisor. The local government should set reliability targets; for example, 75 percent of the inspector's ratings could be required to agree with those of the supervisor's, and 90 percent to agree within one-half point. If the checks do not indicate sufficient accuracy, immediate retraining of raters and possibly further refinement of photos and written standards for the ratings should be undertaken.

Is Management's Evidence Feasible To Develop?

The question of information feasibility is heavily, although not exclusively, a cost issue. Sometimes, other issues, such as confidentiality, can overwhelm basic cost.

Rule To Follow

Management's evidence is feasible to develop when the cost of the information system falls within reasonable operating and evaluation budgets and the information is politically acceptable to develop.

Procedure To Follow

Tables 14 and 15 are illustrative checklists to use in examining the financial feasibility of information systems

The basic procedure is to:

- Estimate the costs of collecting, storing, processing, and verifying the information for each event.
- Calculate rough system costs in terms of staff (internal) and dollars (contract cost)
- Estimate the cost burden on program operator staff (if they are required to supply data) Verify with operators.
- Review system data collection process with those expected to supply data (generally program operators). Determine relative acceptability of data collection scheme.
- Communicate findings to management to determine whether total burden (including operator costs) is acceptable

TABLE 14 Sample Checklist for Determining Costs Associated with Running a Data System

	Cost of Data System	
	Contract	Staff
Data Collection		
Storage		
Processing (by) • • • • •		
Quality Check		

TABLE 15 Sample Checklist for Determining Cost Feasibility

Event or Description	Availability of Required Measurement System (Yes/No)	Measurement System Cost $ staff time or unknown		Measurement Method for Checking Errors (Yes/No)
		Cost	Total Budget for Event	
A. _____				
B. _____				
C. _____				
D. _____				
E _____				
Total Cost				

The cost of information systems is almost inevitably a subject of controversy in government. The OMB forms clearance function was developed expressly to provide a check on information proliferation. Often, the only cost included in analyses of systems is that incurred in developing a system under a contract, while costs of implementing the system, checking its quality, or paying program operators are generally not included. Frequently, operator systems, already in place, are ignored, thus increasing the expected cost of the system because of required modifications to operator systems. The price of ignoring such costs is that management can be led to make decisions they will later have to reverse.

Example

An example in which all factors were not accounted for is provided by the Legal Services Corporation (LSC). LSC and its progenitor, OEO's Office of Legal Services, had been criticized for some time by the General Accounting Office for the failure to implement a basic information system. LSC was in the process of mounting a series of demonstrations to assess alternative legal services delivery methods, and one part of the effort involved the collection of cost and workload data. Under contract, a project reporting system was developed to collect the information required by the evaluation design. The system involved adoption of uniform case file forms in each office, the use of unique client identifying numbers, and the recording of time invested by individual attorneys.

The system relied on the collection of disaggregated case-related data being forwarded to headquarters for processing and analysis. It imposed costs on individual attorneys, projects, and the corporation, and added costs to monitor reliability. Having developed the system, however, LSC was then unable to implement it. Important among the many issues raised was the subject of potential or feared violation of client confidentiality pledges. Further, project directors believed the system intruded too far into their domain, threatening their autonomy.

Even if the actual financial cost had proven acceptable to management, a possibility because of an offer of financial assistance made to LSC, the political costs proved to be unacceptable. A compromise system had to be and was developed.

It should be noted here that there exist few usable ground rules for determining reasonable costs. The so called 1-percent policy, in which the Secretary of DHEW is authorized by Congress to spend up to 1 percent of program budgets for evaluation, is not a guide. There are instances in which over 50 percent of a demonstration program budget was spent for evaluation, and others in which 1 percent would be high. In fact, there is no substitute for costing out the whole information system, even quite roughly, and allowing management to determine whether the cost is too high.

Are Management's Expectations Plausible Given Their Intended Use of Information?

Management's expectations derive, in part, from the basic design and assumptions of the program and in part from their presumed ability to redirect the program based on evidence of performance.

Rules To Follow

Management's expectations are plausible when there is evidence that deviations from expected performance can be identified by management and that management can affect program activities to correct for performance deviations. Table 16 provides a checklist to use in reaching conclusions about management's use of information as a feedback device to correct performance deviations.

Procedure To Follow

The procedure to follow in this case requires the evaluator to interview management about planned use.

TABLE 16 Checklist to Assess Planned Use of Information

Assessment	Event 1	Event 2	Event 3	Event 4
1. There is evidence that management has defined acceptable performance levels (Yes/No).				
2. There is evidence that management has onsite activities (or plans) to correct performance when it deviates from acceptable performance levels.				
3. Management planned use is plausible (Yes/No).				

- Interview management to obtain a definition of the performance levels that would trigger either direct action or further inquiry, as based on a description of expected values for key events, and a definition of the existing or planned activities controlled by management which are believed capable of affecting the performance of the event(s) in question.

- Analyze the functional description of the program and attempt to link management's planned actions to functional elements which govern the event performance.

- Note any gaps between management's planned actions and the event in question.

- Form judgment as to the plausibility of management's intended use of information, and record them on a checklist as per Table 16.

- Communicate findings to management. Continue to work with management until a plausible management event sequence is developed to affect performance.

In practice, the evaluator is being asked to extract from management another program description—this one concerned with the indirect intervention. The program events must, again, begin with resources and staff activities over which management has control and end with the event(s) whose performance is to be corrected.

This last analysis is, if not ignored, at best completed halfheartedly in most evaluations. The problem seems to stem from at least two causes. First, managers often do not know exactly how they intend to use information and, thus, cannot tell evaluators. Second, there continues to be a rather generalized belief that performance information will automatically tell management how to correct problems.

This latter point is especially problematic. It may follow from the general notion that evaluation is a method of assessing experiments and that findings are supposed to be unequivocal. Yet few Federal programs are experiments in any way other than name alone, and most evaluation contracts ask the contractor to produce findings and recommend solutions to observed problems. It is our general contention, to the contrary, that solutions must come from those responsible for managing programs. Further, it is our belief that at least the outlines for plausible solutions should be developed before rather than after an evaluation study is completed.

Example

We cited earlier the example of a performance signaling system developed for the Atlanta public school system which failed because the evaluators did not define the real functions carried out by school administrators and resource specialists. In other words, had we known the real program activities, we would have said that the program expectations were implausible.

This same study serves as an example of the issue being discussed here. The planned management use of the information could not have affected performance as intended because management had not designed its indirect intervention to use such information. Instead, use was left relatively undefined. It was *assumed* that performance information would trigger allocation of resource teachers, but it did not. It probably was assumed, too, that technology existed (teaching methods known to be effective) which would correct performance deficiencies, but it did not. What was required here was, as already noted, a fundamentally different type of indirect intervention—one geared to research and development—in which performance deviations were only the first step in a

multistep, corrective process involving the study of successful methods and the development of technology and methods to use it.

The earlier cited Office of Legal Services quality assessment method, on the other hand, was plausible, precisely because management had carefully designed its planned use of information in such a way that the actions were plausible. OEO, in fact, used the system as planned because its planned use already matched existing functions. The information system merely formalized a process already in place

The point is planned use cannot be treated lightly, a fact that is amply demonstrated by even casual reading of most evaluation plans prepared by agencies. We close this subject with material obtained from a prior study of evaluation planning (Horst et al., 1974a).

"The FY 1974 (DHEW Evaluation Planning) guidance asked the program offices to specify the issues which evaluation should address." This was done to help clarify exactly how evaluation was to be used However issues in most submissions were portrayed in such general terms that it was hard to tie them to any particular decision context Typical issue statements included: "Determination of overall effectiveness of various types of services in relation to cost under various conditions." "development of system management approaches to improve services," and "the relative effectiveness of current and proposed mechanisms for delivery of services in the public and private sectors."

The guidance also asked respondents to specify users and to associate users and issues in order of priority We found that most program audiences defined almost everyone in the world as potential users of their evaluation studies For example, one program office specified the Congress, the Secretary of DHEW Health Services and Mental Health Administration, the National Institute of Mental Health the program director, regional offices, States, local project directors, and other government and private agencies. No priority among them was indicated. To illustrate how another program office described evaluation users, their written responses are presented in schematic form in Figure 12 The boxes represent users at the Federal, State, and local levels. The arrows show decision processes or products that evaluation was supposed to affect and the groups that would be affected

The use of individual evaluation project findings was treated in general terms, as the following quotations from plan submissions illustrate

- knowledge to assist NIMH, States, and facilities to plan useful data systems
- technical assistance and planning
- guidance to State and county governments and increased NIMH and Regional Office evaluative capability
- Results of this study will assist planners and project directors to redirect their programs for better achievement of their goals and objectives

To be used evaluations must be treated as new products for whose use substantial design engineering work is required It cannot be left to chance

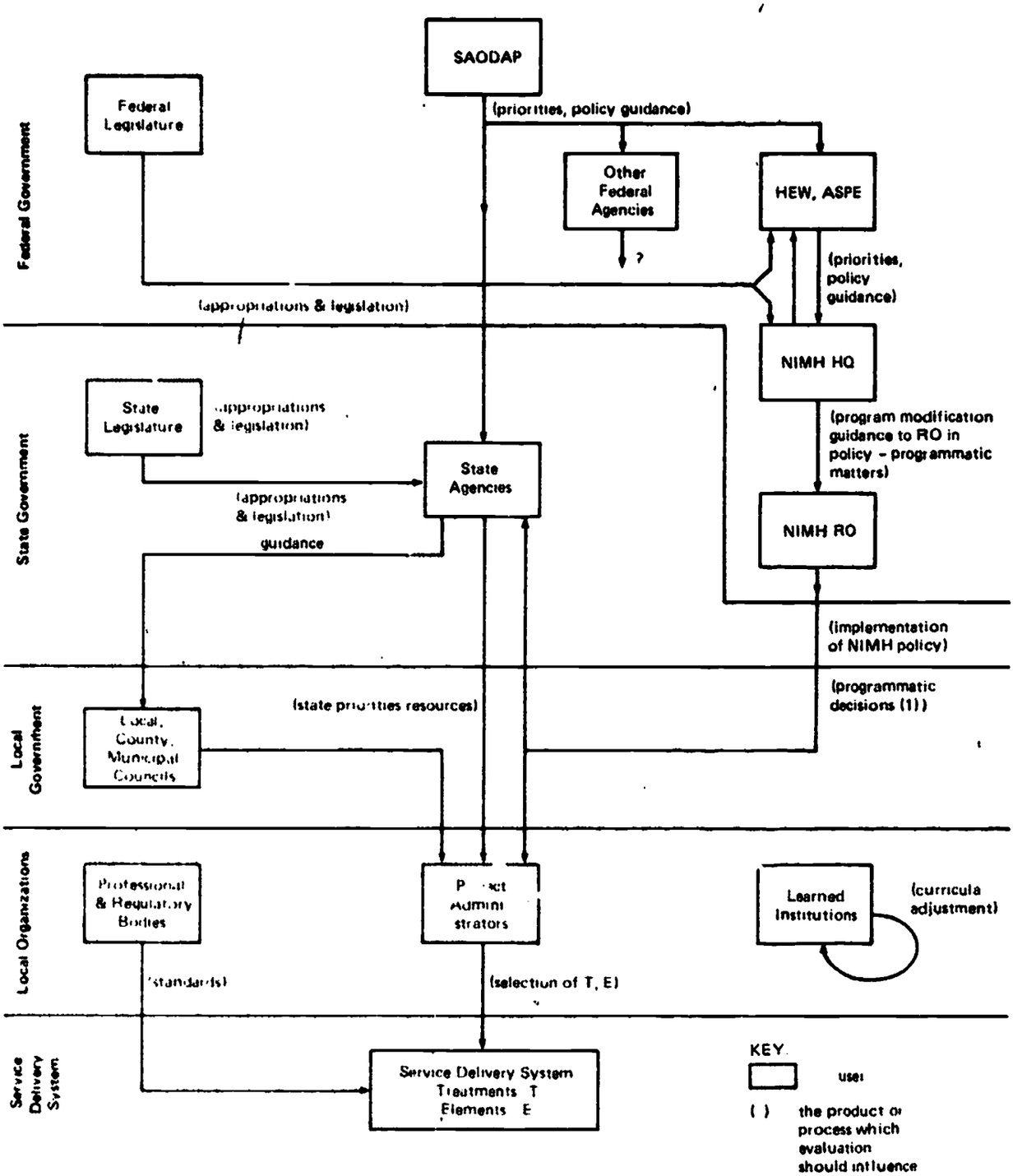


FIGURE 12 Schematic of Users and Uses of Evaluation Based on the Division of Narcotic Addiction and Drug Abuse, FY 1974 Evaluation Plan Submission.

IV. THE THIRD STEP: CONSTRUCT ALTERNATIVES

Early applications of EA led to an unsatisfying conclusion—the one or two potentially evaluable descriptions concerned studies of relatively low interest to management, while more interesting studies were blocked by the EA conclusions which suggested the need for further definition or actual program design changes. It seems, then, that EA could prove to be a highly effective brake on expenditures for evaluation. This particular outcome, while somewhat appealing, does not take into account that the evaluator, having created an apparent impasse, should be the one to define alternatives which are helpful to management. It is probably also true that an evaluator's conclusions regarding problems in program descriptions could be strengthened by defining and offering management ways to resolve the observed problems.

We define, then, an added task for the evaluator doing an EA—construct alternatives wherever possible These can take the form of alternative expectations, program activites, or information approaches For any of the threats to program achievement named throughout, evaluators can suggest one or more of these alternatives to fill gaps in the program logic. Often, the alternative might be a developmental process of some kind to discover and define an appropriate alternative.

Alternative Expectations

Perhaps the most common problem affecting Federal programs is unrealistic expectations, most likely resulting from a management decision which is invalid, implausible, or unacceptable to policymakers.

For any one of these problems, one possible solution is altered expectations. Most Federal programs are defined legislatively with a specific budget level and a relatively limited definition of the specific activities that will or must be present to fulfill the intent of the legislative and executive branches Often, the EA will be the first time anyone seriously examines the underlying logic of the design and the implied logic of the real program as it is being operated in the field.

Alternative expectations can be either different or lowered ones. For example, many programs are expected to achieve more than one objective, and the EA may reveal that two of six objectives are implausible, given the nature or size of the program. An option here is to simply drop the two believed to be implausible Assuming that the two implausible objectives are on the low-end of the priority scale, they could be dropped without endangering the underlying political support for the program. Basically this is the issue to be explored by the evaluator. Program support is often provided within certain performance limits, one end indicating success and the other failure, and the range between the two may be wider than imagined.

One example of the altering of expectations is available from work completed for the Appalachian Regional Commission. An EA of the Commission's health and child development program revealed that economic impact was one of many objectives defined for the program. The term had never been defined, nor had any system ever been developed to collect such information. Clearly any investment of Federal funds in a region creates an economic effect in staff salaries and purchase of goods and

services However, examination of the legislation and its background revealed that the legislators inserted health and child development into the act primarily for social and humanitarian reasons, rather than solely economic ones In short, the health and child development program would not be considered unsuccessful were its economic impact limited.

Commission policymakers were apprised of the results of the analysis and became convinced that it would not be worth the investment of additional resources to further define the term and establish a system to collect evidence about the economic effects of investments in health and child care. The expectation, accordingly was dropped.

Alternative Program Activities

In some cases, policymakers or managers will have expectations about which, however implausible they feel strongly As one might expect, the evaluator has a more difficult task ahead of him when this is the case Generally, when faced with an objective which is unsupported by real activity, three possibilities exist experts can be brought in to design a program that will work; program operators who are successful can be identified, and their methods documented and disseminated; or a research or demonstration program can be undertaken to develop the knowledge to design a workable program

Which of these approaches is the most appropriate depends on the state of knowledge about the objective and the likelihood that success in this area is being experienced somewhere in the program (i e , some local program operators already know how to achieve the objective).

One of the more common approaches used in government is to form a task force of people knowledgeable in the field to design a program. Less often is it recognized that local program operators, through necessity, may have already produced the answer to what works. The selection of an approach depends, then, on how much information is already available about local programs and successes

Within DHEW, the approach being taken is to identify locally successful programs via telephone surveys and selected site visits. Telephone surveys are designed to identify plausible candidates; site visits are intended to document and validate the claims of the local HSA directors.

Whichever approach is adopted, the intent is to provide program management with a way out of the dilemma created by an EA finding which shows the objective to be implausible. Management can more easily act when presented with alternatives, than when it is told simply, "Your program won't work," or "It can't be evaluated."

Alternative Information Approaches

Often expectations, however ambiguously defined, or program activities cannot or will not be altered The evaluator must then design an information approach which can, by successively acquiring information, refine the manager's understanding of what is possible, and, thus, slowly define that part of the program description which is ambiguous.

One such approach was developed for the Law Enforcement Assistance Administration—the National Evaluation Program (Nay et al , 1977) LEAA, as earlier noted, funds hundreds of different types of local projects Many of these projects, although categorized under one name, operate very differently from locale to locale The approach taken in one such project was to purchase information in stages, each representing a potential stopping point.

LEAA and State staff carried out an initial stage in which project groupings—topic areas of potential interest to Federal and State administrators—were identified. A topic area was a group of projects presumed to be focused on a similar objective and with roughly similar methods.

Phase 1 studies were then contracted out. These studies included:

- A literature search, development of a general taxonomy of projects, and an assessment of what was known about the topic area;
- A description of a sample of field projects, including the development of models;
- A generalized analytical framework which would represent the projects reviewed during onsite work;
- An assessment of the present state of knowledge, including an analysis of the accuracy and reliability of current information;
- A design for a Phase II evaluation; and
- A design for a single project in this class which could be used by State or local officials.

With this information gathered, LEAA could stop, because they believed they had enough information, or proceed to Phase II, a complete evaluation based on the Phase I design. This approach enabled LEAA to gather comparable information about an extremely diverse set of projects.

Another example of using an information approach to clarify strongly held but ambiguous objectives can be found in the study completed for the Appalachian Regional Commission. One of the purposes of Congress in creating ARC was to provide a vehicle for developing and demonstrating unique, innovative solutions to the region's problems. The term "innovation" was used liberally in the legislation and internal planning documents, but had never been defined. On further examination, the objective seemed to be taken seriously by policymakers, yet not only was it not defined well, there were also no functional activities in place to achieve this objective. Defining "innovation" in the abstract was possible, but was not considered useful. Defining the term operationally was attempted, but it proved to be elusive. Yet, when queried, staff and management seemed able to cite examples of innovative projects and to reject some projects as not innovative.

Accordingly, it was decided to establish a rather loosely defined process in which a particular project would be declared innovative if local, State, and Commission staff could mutually agree that it was so, and why. The process involved screening of all project grant applications at the three levels, agreeing on a set of innovative projects, selecting the most promising ones (based on definition of objectives and method in the grant application), and then evaluating those projects. The evaluations would be used to promote and [illegible]pt to replicate successful innovations, and to advocate expanded financial support for pr[illegible]echniques. Here an information process was being used to define, through example and study, an otherwise highly ambiguous objective.

V. THE FINAL STEP: IMPLEMENT EVALUABILITY ASSESSMENT FINDINGS

The completion of an evaluability assessment is not the point at which the evaluator documents and presents his findings to management. The final step occurs when management makes the three sets of decisions necessary to conclude the assessment. Having decided to accept or reject the evaluator's conclusions, chosen which evaluable description (current or alternative) to implement, and determined how to organize and staff the effort to implement its decisions, the evaluator must then assume an important role in implementation.

Management Decisionmaking

The evaluator's major tasks in the decision process are to organize the decisions to be made such that management can understand the options available, and then record the decisions.

In the EA completed for the Bureau of Health Planning, a formal report was prepared to lay out the options in a form that would facilitate the decision process. This paper (Wholey et al., 1977) was distributed to the Bureau's management group—the director and his immediate supervising staff.

Information was presented both in narrative form and in several sets of tables, one of which communicated the options available to management, another of which was provided blank to enable the management group to record their reactions to the EA findings, and a third of which was provided blank to enable the management group to record their decisions on which options to pursue.

Tables 17 through 22 are the first set, those summarizing the options available to management. Tables 17, 18, and 19 are organized around the options concerning three of the Bureau's programs. (The Act, P.L. 93-641, was divided into a number of programs for purposes of the evaluability assessment.) Table 20 summarizes the types of evaluation data and the sources of those data by evaluation option. Table 21 was prepared to summarize for management the implications of selecting certain options or sets of options relative to Bureau management objectives. For example, if Bureau management wished to focus on increasing the effectiveness of its technical assistance effort, options were presented to facilitate that objective. Finally, Table 22 was presented to summarize the options which could be adopted to support different management purposes.

The key here was to get management to rank order choices. It was clear that there were not adequate funds to pursue all options. So choices had to be made. The tables and accompanying text were designed to aid in that task.

Tables 23 through 28 summarize the reactions to findings: Tables 23, 25, and 27 summarize reactions to findings about the three programs—the extent to which the programs were ready to be evaluated, while Tables 24, 26, and 28 then summarize the reactions of the management group to the evaluation/management options developed by the evaluators. These reactions preceded discussion and decisions relative to ranking the options in order of importance to the Bureau.

TABLE 17 Evaluation/Management Options: Information Sources, Costs, and Uses of Information (Bureau Program I)

Options	Evaluation Information	Information Source	Approximate Annual Cost		Use of Information
			$	Staff	
I.1 Manage Program I to support DHEW production of regulations and guidelines	(1) Cost to Bureau of drafting regulations, guidelines, and reporting systems	MITS and other Bureau management systems (See p. 31[1])	40,000 - 80,000	0.25 man-years	Managing Bureau activities, influencing those above the Bureau
	(2) HSA/State Agency/ Regional office reactions to DHEW products	Telephone surveys (See pp. 32-33[1])			
I.2 Redesign Program I to minimize DHEW production costs and time delays	Cost of Bureau drafts of regulations, guidelines, and reporting systems	MITS and other Bureau management systems (See p. 31[1])	0	Minimal	Obtaining changes in the DHEW/Bureau process for developing and clearing required products
I.3 Redesign Program I to minimize the negative effects of DHEW regulations and guidelines on agencies	Likely impact of proposed regulations on agency resource allocations and performance	Site visits to 5 to 10 agencies (See pp. 36-37[1])	70,000[3]	Modest	Recommending changes in proposed regulations
I.4 Redesign Program I to support DHEW development of realistic federal objectives	Evaluation information from Options II.1/II.2/II.3 or from Option III.4	Information from Options II.1/II.2/II.3 or III.4 (See tables following)	Same as Options II.1/II.2/II.3 or III.4		Recommend to DHEW realistic Federal objectives for HSAs and State agencies

[1] These page references are to *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Bureau Program I: Rulemaking*, Urban Institute Contract Report 5906-20-02, November 1977.

[2] $5,000 to $10,000 per product.

[3] $70,000 per evaluation

TABLE 18 Evaluation/Management Options: Information Sources, Costs, and Uses of Information (Bureau Program II)

Options	Evaluation Information	Information Source	Approximate Annual Cost			Use of Information
			$		Staff	
II.1 Manage Program II to have HSA and State agency planning produce evaluable programs	Agency progress in developing realistic objectives and plausible programs that will achieve those objectives (See pp. 61-65)[1]	Health systems plans; annual implementation plans; State health plans; State medical facilities plans Regional office site visits to obtain data confirming program plausibility	0 or 70,000 280,000	& & 	2 to 5 man-years 1 man-year	(1) Reporting to DHEW and Congress on nationwide performance in health planning and expected results over the next 5 years (2) Policy guidance to regional offices, HSAs, and State agencies (3) Identifying outstanding agencies and problem agencies
II.2 Manage Program II to have HSA's and State agencies achieve federally defined objectives	Agency progress toward national guideline standards or other federally defined objectives (See pp. 70-71)[1]	Health systems plans; annual implementation plans; State health plans; State medical facilities plans	0 or 70,000 140,000	& & 	2 to 3 man-years 1 man-year	(1) Reporting to DHEW and Congress (2) Policy guidance to regional offices, HSAs, and State agencies (3) Identifying outstanding agencies and problem agencies.

(continued)

TABLE 18 Evaluation/Management Options: Information Sources, Costs, and Uses of Information (Bureau Program II) (Continued)

Options	Evaluation Information	Information Source	Approximate Annual Cost			Use of Information
			$		Staff	
II.3 Manage Program II to have HSAs and State agencies achieve local and State objectives	Agency progress toward local and State objectives (See pp. 61-65)[1]	Health systems plans; annual implementation plans; State health plans; State medical facilities plans	0 or 70,000 280,000	 &	2 to 5 man-years 1 man-year	(1) Reporting to DHEW and Congress (2) Policy guidance to regional offices, HSAs, and State agencies (3) Identifying outstanding agencies and problem agencies
II.4 Manage Program II to have HSA's and State agencies establish required functions, structures, and operations	Not examined	Not examined	Not examined			Not examined

(continued)

TABLE 18 Evaluation/Management Options: Information Sources, Costs, and Uses of Information (Bureau Program II) (Continued)

Options	Evaluation Information	Information Source	Approximate Annual Cost		Use of Information
			$	Staff	
II.5 Redesign Program II to develop routinely HSA/ State agency performance standards	See Option III.4 (See tables following)	Information from Option III.4 (See tables following)	Same as Option III.4 (See tables following)		Use information on HSA and State agency performance to develop agency performance standards
II.6 Redesign Program II to estimate routinely the effects of congressional/ DHEW/Bureau policies on agency resource allocation and performance	Agency resource allocation (See pp. 77-79)[1]	Agency work plans, or agency work plans plus continuous monitoring of 5 to 10 agencies	0 to 200,000	& 2 to 5 man-years[2]	Recommend changes in regulations and guidelines
			or		
			70,000-480,000	& 1 man-year	

[1]These page references are to *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW Bureau Program II: Administering the Law Nationwide*, Urban Institute Contract Report 5906-20-03, November 1977.

[2]The same staff could accomplish evaluations II.1, II.2, II.3, and II.6.

TABLE 19 Evaluation/Management Options: Information Sources, Costs, and Uses of Information (Bureau Program III)

Options	Evaluation Information	Information Source	Approximate Actual Cost		Use of Information
			$	Staff	
III.1 Manage Program III to complete products on schedule	Planned vs. actual schedules of activities and products (See p. 32)[1]	MITS and other Bureau management systems	0	Mini-mal	Allocation of Bureau staff and funds to production of technical assistance products
III.2 Manage Program III to produce products considered useful by agencies	HSA/State agency reactions to Bureau technical assistance products (See pp. 32, 34 & 35)[1]	Telephone surveys	50,000 - 400,000[2] &	½ to 3 man-years	Allocation of Bureau staff and funds to production/revision of technical assistance products
III.3 Manage Program III to produce products which are used and are effective	HSA/State agency use of Bureau technical assistance products (See pp. 32, 39 & 40)[1]	Plans, other agency documentation	100,000 - 800,000[3] &	1 to 3 man-years	Allocation of Bureau funds to production/revision of technical assistance products
III.4 Redesign Program III to increase the use and effectiveness of discretionary technical assistance	(1) Processes and products through which effective HSA's and State agencies have defined problems, reached solutions, and brought about changes in specific health services (2) Changes in the health care system (See p. 52)[1]	(1) Telephone surveys (2) Site visits to 8-10 HSA's/State agencies	1,000,000 - 2,000,000[4] &	5 to 10 man-years	(1) Technical assistance to HSA's and State agencies (2) Support for Bureau Programs I, II, and IV

[1] These pages references are to *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Bureau Program III: Developing and Disseminating Technical Assistance Products* Urban Institute Contract Report 5906-20-04, November 1977.

[2] $5,000 to $10,000 per product.

[3] $10,000 to $20,000 per product

[4] $50,000 per evaluation

TABLE 20 Types and Sources of Evaluation Data

Type of Evaluation Information	Source of Data	Options
1. Bureau costs and progress in producing required and discretionary products	A. MITS and other Bureau management systems	I.1, I.2, III.1
2. Opinions of HSA, State agency, and regional office staffs as to the usefulness of Bureau products	B. Telephone surveys of HSA, State agency, and regional office staffs	I.1, III.2
3. Use of Bureau products in HSA/State agency activities	C. HSA/State agency plans and other agency documentation	III.3
4. HSA/State agency progress in developing plausible plans and realistic programs to affect the health care system	C. HSA/State agency plans[1] and other agency documentation	I.4,[2] II.1
5. HSA/State agency progress toward national guideline objectives and other local, State, and national objectives	C. HSA/State agency plans[1]	I.4,[2] II.2, II.3
6. HSA/State agency resource allocation	C and D. HSA/State agency work programs and site visits to agencies	I.3, II.6
7. Effective HSA/State agency problem solving and resulting changes in the local/State health care system	D. Site visits to HSA's/State agencies	I.4,[2] II.5,[3] III.4

[1] New planning guidance would be required, establishing new formats for presenting health systems plans, annual implementation plans, State health plans, and State medical facilities plans.

[2] Option I.4 uses information from Option II.1/II.2/II.3 or III.4.

[3] Option II.5 uses information from Option III.4.

TABLE 21 Evaluation/Management Options and Bureau Management Objectives

Program I: Rulemaking	
Current	(No option described)
Stronger internal bureau management	Options I.1 and I.2
Stronger bureau role in national policy development	Options I.3 and I.4[1]
Program II: Administering the Law Nationwide	
Current	(No option described)
Stronger internal bureau management	Options II.1 and II.2
Stronger bureau role in managing a nationwide program	Options II.1, II.2, II.3 and II.5[1]
Stronger bureau role in national policy development	Option II.6
Program III: Developing and Disseminating Discretionary Technical Assistance Products	
Current	Option III.1
Stronger internal bureau management	Option III.2
Stronger bureau technical assistance role	Options III.3 and III.4

[1] Options I.4 and II.5 are Bureau staff activities which use evaluation information from other options. Thus, they cannot be implemented alone

TABLE 22 Alternative Decision Packages (Combination of Evaluation/Management Options)

Evaluation/ Management Option	Package 1: Current Operations	Package 2: Stronger Bureau Management of Programs I, II, and III	Package 3: Stronger Bureau Technical Assistance Role	Package 4: Stronger Bureau Role in Management of a Nationwide Program and in National Policy Development	Package 5: Stronger Bureau Role in Technical Assistance and in National Policy Development	Package 6: Stronger Bureau Role in ● Management of a Nationwide Program ● Technical Assistance ● National Policy Development
I.1 (Support DHEW)		X	X	X	X	X
I.2 (Minimize cost/delay)				X	X	X
I.3 (Minimize negative effects)				X	X	X
I.4 (Recommend Federal objectives)				X	X	X
II.1 (Evaluable programs)		X		X		X
II.2 (Federal objectives)		X	X	X	X	X
II.3 (State/local objectives)				X		X
II.4 (Required functions)			Not Examined			
II.5 (Develop standards)						X
II.6 (Federal effects on agencies)				X		
III.1 (Products)	X					
III.2 (Useful)		X		X		
III.3 (Use)			X		X	X
III.4 (Models of agency problemsolving)			X		X	X

TABLE 23 Summary of Bureau Managers' Reactions to Findings on Program I: Rulemaking*

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More Evidence is Required to Convince Me	
1. Program I is well-defined in terms of agreed upon, measureble objectives.	9 "A good half of the problem is how ably we manipulate the system . . . We don't do that very well."	0		
2. The Bureau's program has not met its objectives in the past, as measured by available data sources. Already available information indicates that Program I does not complete required products on the time schedules required to meet the needs of State and local health planners	9	0		
3. The Bureau's program is perceived as not working by Bureau management and by those above the Bureau.	7	2 "Don't agree now . . . A few months ago I would have . . . It is plausible now. . . . more vertical linkages." "Foley may feel urgency, Endicutt never did."		The Corbett-Foley Committee may be helping.

(continued)

TABLE 23 Summary of Bureau Managers' Reactions to Findings on Program I: Rulemaking (Continued)

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More evidence would be required to convince me	
4. An examination of the program indicates that Bureau's program logic is not plausible.	5 "Internal constreints may be even more severe than external."	4 "Half right . . . The HEW environment is half the problem; the Bureau is the other half." "New Administration policy."		
5. The program may have serious negative effects on HSA/State agency operations, imposing unrealistic requirements on HSAs and State agencies, and causing costly disruptions in State/local operations.	3	2 "What is produced is generally good and problems are resolved quickly."	4 "Philosophically I am disposed to believe it, but I haven't seen the evidence . . . Some agencies are doing well in spite of me." "Agree with 'unrealistic' . . . More evidence required re 'costly, disruptive.' " "Anyone outside of Federal government would generally react negatively to regulations and guidelines." "I don't understand the regulations and guidelines re Plans myself . . . On the other hand, the Law is very complicated."	More evidence could be presented, to clarify the findings and demonstrate its correctness.

*The large numerals indicate the number of Bureau managers who agreed or disagreed with the findings, or needed more evidence to be convinced.

TABLE 24 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program I (Rulemaking)

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1, 2, . . ., 14)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? ($/Year)
I.1 Manage Program I to support DHEW production of regulations and guidelines	7 Yes 1 - Not fully	5 – Yes 1 – Yes for guidelines; no for regulations 1 - Yes if e reporting system existed (not too sanguine about this)				staff + staff +	15 - 40,000 40 - 50,000 40 - 80,000 Start with 1 and see what we get 2 - 3/year (all internally developed guidelines)
I.2 Redesign Program I to minimize DHEW production costs and time delays	7 Yes 1 - Not fully	4 – Yes 1 - Yes for guidelines; no for regulations 1 - no: not feasible to minimize HEW costs and time delays unless someone at top . . . Champion giving us help . . .					

(continued)

TABLE 24 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program I (Rulemaking) (Continued)

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1,2, . . ., 14)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? ($/Year)
I.3 Redesign Program I to minimize the negative effects of DHEW regulations and guidelines on agencies	7 – Yes	3 – Yes 1 – Yes for guidelines; no for regulations 1 – No 1 – Probably hard to do 1 – Limited regardless of what we do		"Your thinking is good but has been overtaken by events."		Less than 1 man-year	$100,000 + $1[illegible]00 All [illegible]n spare
I.4 Redesign Program I to support DHEW development of realistic federal objectives	7 – Yes	6 – Yes		"Looks pretty expensive, probably with questionable chance of success."		3 man-years +	suitable amount of contract $

(continued)

TABLE 24 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program I (Rulemaking) (Continued)

Options	Why should the Bureau commit resources to this option?	Given the commitments of your Office/Division would you commit your own resources to this option?				Where should the necessary resources come from?		
		(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-year/Years)	How Much Contract Money? ($/Year)	Management Time	Staff Time	Contract Money
I.1 Manage Program I to support DHEW production of regulations and guidelines.	"Creative discipline." "Good chance of benefit."			2-3 months/product	$5-10,000 to start		X X X	1% evaluation $ or own staff X
I.2 Redesign Program I to minimize DHEW production costs and time delays				2-3 months/product			X X	X
I.3 Redesign Program I to minimize the negative effects of DHEW regulations and guidelines on agencies	"NPRM is another way to get information from the field. A lot of agencies are doing very thorough analyses."						X	Ise travel $ not used correctly
I.4 Redesign Program I to support DHEW development of realistic federal objectives	"It's the most positive approach. It makes the most logical sense." "Need for *ex post* analysis and adjustments."			DPMT: "I couldn't guess. I would be willing to commit the resources that would be necessary."				

TABLE 25 Summary of Bureau Managers' Reactions to Findings on Program II: Administering the Law Nationwide*

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More evidence required	
1 The current state of program definition falls short of that intended by the Bureau and required by those above the Bureau. Much of the intended program (e.g., Event 6a, "Achieve State/Local Objectives," and Event 7, "Achieve National Objectives") has not been defined in measurable terms.)	9 "The requested adjustments [exceptions] to the national guidelines will give some indication of what State and local objectives are."	0		
2. While monitoring of progress toward the objectives in the proposed national guidelines, Event 6b, is probably feasible, it is by no means clear that HSA/State agency activities are likely to achieve these objectives.	8 " 'Not clear' is about the softest way you could say could say that " "I don't want you to consider that we've done something wrong . The issue is in doubt until the battle is fought." "The national guidelines really are not applicable to HSAs because they are not health service market areas." "There is a lack of basic skill . . to insure that HSAs could prevail and achieve objectives . . . "	1 "Don't agree completely. We will move into strong exceptions. That will have to be built into the picture . . . I think on some national guidelines you might see some progress."		

(continued)

TABLE 25 Summary of Bureau Managers' Reactions to Findings on Program II: Administering the Law Nationwide (Continued)

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More Evidence Required	
3a. The Bureau's Program 11 does not appear to be producing realistic, plausible performance standards. The Bureau does not have either the information collection process or the inhouse staff activities necessary to set realistic standards for HSA/State agency performance.	6 "Certainly, in terms of outcome, I agree." "There are constraints in the law that get us into the problem." "Too few internal resources committed . . Limited interaction with people in the field Far too much emphasis on anecdotal syndrome . ""	3 "We're awfully slow and late . . . We will be producing realistic performance standards . . . The information can only come after we adopt some standards, do some site visits, and see what information we get . . . " "In some respects, yes; in some, no. We're on the verge of having a fairly adequate process for judging development on HSPs and AIPs." "We can set some performance standards in terms of the Law, but we don't have much of anything to collect information or the ability to examine it . . . We don't have information to set standards on results what changes in health status or quality you can reasonably expect and what measures."	"[After briefing:] You've convinced me."	The finding should be more clearly stated. The finding uses the term "performance" to mean outcome or impact, not process.

(continued)

TABLE 25 Summary of Bureau Managers' Reactions to Findings on Program II: Administering the Law Nationwide (Continued)

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More Evidence Required	
3b. Moreover, the proposed reporting system seems suspect on cost feasibility and cost-utility grounds.	6 "We are a long way from knowing how we will use the data." "We haven't established the basis for the information we need." "No unified position on how the information would be used."	2 "We have to do it . . . I don't know that it is very expensive." "What is suspect is our ability to display the data . . . "I think it's worth the cost because the information we'll be getting is the bulk of the information we need to defend the program in Congress . . . It puts us in position to show that we at least know what's going on in the agencies."	1 "I don't know . . . Agree in general . . . Your third finding I support." "Not convinced."	
4. Some of the congressional/DHEW/Bureau requirements placed on agencies are resulting in unrealistic plans and programs, dissipation of agency resources on unproductive activities, and a loss of agency credibility in planning. (The law and Federal regulations and guidelines require HSAs and State agencies to plan for achieving objectives over which they have no control, for example, to improve health status, environmental health, and occupational health.)	5 "This is a very basic issue." "Agree to a certain extent . . . The emphasis should be on congressional requirements."	3 "Mostly disagree . . . We do talk about health status goals . . . Those agencies that want to pick up on that [do]; those that don't, don't . . . We aren't forcing them to deal with these . . . They aren't so passive. If [the requirements are] unrealistic . . ., they call RO [and us] . . ." "Disagree . . . But I don't think anyone takes them seriously." "Unrealistic plans evolve more from luck to prototype [plans] . . . We're not even sure what goes into a HSP."	1 "Philosophically I believe it, but I haven't seen the evidence . . . Some agencies are doing well in spite of us."	More evidence is required to clarify the finding and to demonstrate that it is correct.

(continued)

TABLE 25 Summary of Bureau Managers' Reactions to Findings on Program II: Administering the Law Nationwide (Continued)

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More Evidence Required	
5. HSAs and State agencies are, for selected issues and problems, developing realistic objectives and plausible programs and, in some cases, demonstrating results.	9	0	"I think I'd feel more comfortable if there were more evidence to suggest this."	
6. State and local plans do not presently enable one to distinguish realistic from unrealistic objectives.	5 "Should regional offices have a role here?" "Not only true of State and local health plans, but of almost all planning." "I think we've seen some improvement." "Agree, but we need to have more evidence." "Report doesn't provide that much evidence . . . More evidence is required."	3 "That's a good criticism. It's stated a little too broadly. I'd say that many of them are not sufficiently specific . . . Many are, many aren't . . ." "For the most part, the things we've seen seem to be realistic and not outlandish. . . In many cases, the agencies are kind of conservative, some so much that they're not proposing anything in the community." "In HSP/AIP, regional offices have been able to identify unrealistic and not well defined goal statements and get plans rewritten."	1 "Don.t know . . . Not familiar with the plans."	

(continued)

TABLE 25 Summary of Bureau Managers' Reactions to Findings on Program II: Administering the Law Nationwide (Continued)

Findings	Managers' Reactions to the Finding.			Urban Institute Comments
	Agree	Disagree	More Evidence Required	
7. Though Congress sees the Bureau as having a major role in management of the health planning program, the Bureau's capability to use performance information is not well developed or defined.	8 "We're still ad hoc-ing it from one month to the next." "That's put mildly—beyond the fairly formidable efforts that Pete has made and that's episodic." "There is no clear Bureau policy on what information we need and how we will use it."	1 "True to some extent; not true to some extent . . . HSP's, AIP's: we are using them as a basis of intelligence as to what we should be doing to help . . . and [we] try to take information that demonstrates ability to deal with problems and get it to other agencies."		

*The large numerals indicate the number of Bureau managers who agree or disagreed with the findings, or needed more evidence to be convinced.

TABLE 26 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program II: Administering the Law Nationwide

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1, 2, 3, . . ., 9)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? ($/Year)
II.1 Manage Program II to have HSA and State agency planning produce evaluable programs[1]	8 Yes	6 · Yes 1 · No 1 · We could not impose it except by regulation				1 1+	contract support Urban Institute amount okay . . . Need my decrease through time.
II.2 Manage Program II to have HSA's and State agencies achieve federally defined objectives[1]	7 Yes 1 Wouldn't manage	5 · Yes[2] 1 · Probably 1 · No' I still think the MITS or the site assessments should provide the information 1 Technical problems		"Don't spend resources to extract from the plans."		inhouse + substantial[4] 1	contractor[3] Significant amount of evaluation $ $140,000

(continued)

TABLE 26 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program II: Administering the Law Nationwide (Continued)

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1, 2, 3, . . ., 9)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Men-years/Year)	How Much Contract Money? ($/Year)
II.3 Manage Program II to have HSA's and State agencies achieve local and State objectives	8 - Yes	6 - Yes 1 - Not really		"Probably should rely on anecdotal information."		Substantial 1	$140,000
II.4 Manage Program II to have HSA's and State agencies establish required functions, structures, and operations	7 - Yes	6 Yes					

(continued)

TABLE 26 Summary of Bureau Managers' Reactions to Evaluation/Managment Options for Bureau Program II: Administering the Law Nationwide (Continued)

Options	Why should the Bureau commit resources to this option?	Given the commitments of your Office/Division would you commit your own resources to this option?				Where should the necessary resources come from?		
		(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-year/Years)	How Much Contract Money? ($/Year)	Management Time	Staff Time	Contract Money
II.1 Manage Program II to have HSA and State agency planning produce evaluable programs							X	1% evaluation $ or current UI contract 1% evaluation $[6]
II.2 Manage Program II to have HSAs and State agencies achieve federally defined objectives	" If the whole name of the game " to invest national resources to get things done in terms of federally defined objectives and some of the law " "This is what we are held accountable for "			DHP[5] DRA: "1/8 man-year" DPMT: "We have a couple of man-years in that right now."		The Centers & regional offices could play a role if we could design the right information flow system.	3-4 OPS staff members (?) DHP	1% evaluation $ or current UI contract 1% evalua ion $[6]
II 3 Manage Program II to have HSAs and State agencies achieve local and State objectives								1% evaluation $ or current UI contract 1% evaluation $[6]
II 4 Manage Program II to have HSAs and State agencies establish required functions, structures, and operations				"Might be worth investing in one study."			DHP	

(continued)

TABLE 26 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program II: Administering the Law Nationwide (Continued)

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1, 2, 3 . . ., 9)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? $/year
II.5 Redesign Program II to routinely develop HSA/State agency performance standards	5 - Yes	3 – Yes 1 - No problem doing the study. I question getting all agencies to adopt the standards.	"It's most important that we begin to trumpet some successes."			1-2 +	$200,000 - $250,000 [7]
II.6 Redesign Program II to routinely estimate the effects of congressional/DHEW/Bureau policies on agency resource allocation and performance.	5 Yes	2 - Yes 2 – ? 1 - Yes, but don't work from agency work plans 1 – No	"Not worth the effort since we cannot affect Congress anyway." "The impact may well be on people outside the agency, e.g., the applicant." "I don't know what good it would do . . . It won't keep Congress from adding on."				

[1] "Trying to find the information [in the plans could be a problem]. If we went ahead, should ask for [the information in] some charts at the beginning of the plan."
"HSAs won't have much opportunity to reformat the plans State agencies will call the shots on formats."
"I don't think we should adjust how they plan for the sake of a reporting system."

[2] The exception process [means that] they will have a lot more information at regional offices.

[3] "I'd want my staff involved in this."

[4] "The big problem is having the Federal capacity to do all the analyses."
"In feasibility terms I think it's more likely if we lean toward contractors."

[5] DHP: "I have proposed a new unit of 7 professionals and 2 secretaries to analyze plans . . . (part of Plans Development Branch), and up to 5 more in a complementary section . . ."

[6] OPEL would make a major investment.

[7] "But [the total] will come up to more than we can do . . . We have $200,000 to $300,000 that we can invest and don't even have to ask. I assume that we can get $200,000 to $300,000 more by a good pitch."

TABLE 27 Summary of Bureau Managers' Reactions to Findings on Program III: Developing Discretionary Technical Assistance Products*

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More Evidence Required	
1. The program is currently being managed and evaluated in terms of a limited objective. "produce Bureau products as planned."	7 "But not sure it is achieving even that."	2 "I really don't know . . . In some instances I would agree; in some, disagree . . ." "Partially agree."		Some Bureau managers didn't distinguish between needs assessment and *post hoc* evaluation of the impact of technical assistance products
2. The currently evaluable program appears to fall short of other objectives the Bureau has for the programs, namely, to "produce products considered useful by agencies" and "produce products that are used by agencies and are effective in agency operations."	7 "[But] we've not discussed the TA program and adopted those objectives." "[But] the feedback from the users of the material has been quite supportive of the Centers for Health Planning Our informal advice is that Centers are doing pretty well except and ." "We really don't have a way of finding out yet how effective our presentations . . . . our technical documents [are] . . ."	2 "I really don't know . . . In some instances I would agree; in some, disagree."	I have great uncertainty about the whole TA area . . . I am not convinced that, if we did it well, it would make a significant difference . . ."	

(continued)

TABLE 27 Summary of Bureau Managers' Reactions to Findings on Program III: Developing and Disseminating Discretionary Technical Assistance Products* (Continued)

Findings	Managers' Reactions to the Findings			Urban Institute Comments
	Agree	Disagree	More Evidence Required	
3. Achievement of these more outcome-oriented objectives is not plausible given the current Bureau program for collecting and developing technical assistance products.	5 "But there are some steps to try to improve the situation . . ." "But changes are in progress." "You'll probably have to provide more evidence [for others] . . . No one is highly enthusiastic [or] finding the Centers highly useful." "I don't know what the Bureau program is . . . It's hit or miss."	3 "Products considered useful is a plausible objective . . . We'll improve the present half-assed job . . ." "With less money, we'll have to select more carefully . . . Bureau management will be more involved . . . We'll do much better . . ." "People tell me there is such a marked difference in the planning program: more material produced, better information: "Useful . . . Keep it coming"		

* The large numerals indicate the number of Bureau managers who agreed or disagreed with the findings or needed more evidence.

TABLE 28 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program III: Developing and Disseminating Discretionary Technical Products

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1, 2, 3, 4, 5, 6)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? ($/Year)
III.1 Manage Program III to complete products on schedule	9 -- Yes	7 -- Yes 1 -- Skip					
III.2 Manage Program III to produce products considered useful by agencies	8 -- Yes[1]	5 -- Yes 1 -- Yes, with slight revision of current pro-program 1 -- ? 1 -- not clearcut	You'd probably be disappointed in the results. Therefore why spend $10,000 to confirm it?			Use UI estimate[2] Some Do 5-6 first year and half that number later . . . Tie into the Centers and have them survey	$10,000 $50,000 $50,000 $25,000

(continued)

TABLE 28 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program III: Developing and Disseminating Discretionary Technical Assistance Products (Continued)

Options	Do you understand the option based on the material provided by The Urban Institute? (Yes, No)	Do you think the option is feasible? (Yes, No)	What priority should it have? (1, 2, 3, 4, 5, 6)	Given the Bureau's other commitments, should the Bureau commit resources to this option?			
				(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? ($/Year)
III.3 Manage Program III to produce products which are used and are effective[3]	8 – Yes "I don't have any faith in our ability to define 'effective.'"	4 – Yes 1 – Yes, but OMB forms clearance might be a problem 1 – No. would not be possible to get precise definition of "use" 1 – Not clearcut	"Limit to TA documents in which we invest a lot and which we regard as crucial." "You'd probably be disappointed in the results. Therefore why spend $10,000 to confirm it."			If we did the Center staff, 30–40: if just internal 2–3 easily Some[4] Might bring in Bureau people, who complain that they don't understand . . .	$100,000
III.4 Redesign Program III to increase the use and effectiveness of discretionary technical assistance	8 – Yes (At first this option was not clear to some B managers)	5 – Yes 1 – Theoretically feasible, but implementation would be a problem 1 – Yes, but better to use anecdotal feedback	"I wouldn't put much priority on these options standing by themselves. Anything we do here (TA) should be a by-product of the other options. I particularly like identifying successes and insuring that the knowledge is widely shared."			At least 10 A couple/year 6? Pick a few high-priority target areas	$50,000[5]

(continued)

TABLE 28 Summary of Bureau Managers' Reactions to Evaluation/Management Options for Bureau Program III: Developing and Disseminating Discretionary Technical Assistance Products (Continued)

Options	Why should the Bureau commit resources to this option?	Given the commitments of your Office/Division, would you commit your own resources to this option?				Where should the necessary resources come from?		
		(Yes, No)	How Much Management Time? (Man-years/Year)	How Much Staff Time? (Man-years/Year)	How Much Contract Money? ($/Year)	Management Time	Staff Time	Contract Money
III.1 Manage Program III to complete products on schedule								
III.2 Manage Program III to produce products considered useful by agencies.	"Believe information on product acceptability is worthwhile Would like to be more confident that TA will be useful."						DHP – Plans development branch DHP – Agency assistance branch	1% evaluation $
III.3 Manage Program III to produce products which are used and are effective.							DPMT	TA $ or 1% evaluation $
III.4 Redesign Program III to increase the use and effectiveness of discretionary technical assistance	"Would like to be more confident that TA will be useful."	"Pretty easy to see my cynicism: to try to justify the huge investment we've made in TA. I'd like to be able one day to rationalize and say that it has produced some things that agencies have used and that we can point to models of agencies that are effective."					DMP – Plans development branch DPH – Agency assistance branch	1% evaluation $ 1% evaluation $ 1 TA resources

1. "TA is going down . . ." "I question whether the squeeze is worth the juice. I already see a declining investment in TA." "We need this kind of information. The Federal Reports Act should make more distinction between different classes of respondents (e.g., if we create an entity that is largely responsive to us . . .)."
2. "Do it for any TA document we produce by contract or otherwise, even if only 8 respondents."
3. "Rather than get them to rip something out (existing documentation), I'd ask people to write down for us how they used it."
4. "I was on the Publications Committee. Based on that, I wouldn't do much."
5. "Do at least 1 prototype. [Perhaps cover some of the 9 areas in the proposed National Guidelines and other areas to be added (make people more happy and healthy.).]"

Finally, rankings were assigned by Bureau management to the various options and sets of options presented. To get the rankings, individual meetings were held with management, followed by a group meeting to reach a consensus.

Followthrough

Having recorded management's decisions, it was still necessary to organize the followthrough process. The implementation stage immediately following the EA involves development of evaluation designs—instruments, data collection, and analysis procedures, and collection of preliminary information.

Our experience here suggests the need to maintain management involvement. Accordingly, it is useful to establish a management user group to receive information as it comes in and to continue to provide management guidance to the evaluator.

The Bureau of Health Planning responded to this need by establishing two groups for the implementation process. A policy group was established, comprised of the director, deputy director, chief of evaluation, and the division directors. This group met routinely as the management group for the program, agreeing to meet as often as required to provide guidance. A design group, comprised of members of the evaluation staff, program staff, and contractor staff was also formed. Its job was to design and test instruments and data collection processes which could be implemented subsequently either inhouse or under contract.

The two groups were to maintain the involvement of key Bureau staff and management throughout the implementation phase. Their formation contrasts markedly with other evaluation efforts which are dominated by contractor staff and involve program management and staff only after the fact.

During the implementation phase, as instruments are developed and tested, a substantial amount of information is often developed. It is necessary for the evaluator to continue to provide such information to the user group as it is developed.

The end product of this last stage is a complete evaluation design and the production of preliminary findings gleaned during the design process. Any findings must be clearly labeled relative to the quality of the evidence, especially since they will often be based on extremely limited sample sizes. Nonetheless, such limited information can be useful to management and may obviate the need for full evaluation of certain parts of the program.

VI. SUPPLEMENTARY READINGS

After reading this "how-to-do-it" EA guide, you and your evaluation team will be eager to attempt an actual assessment. In order to perfect the skills necessary for completing this task, you will need to practice them. Speaking and listening, reading and writing, organizing and analyzing are creative processes that are at the heart of a successful EA. The following bibliographic citations have been compiled to aid the professional, researcher, and student in sharpening those skills.

Guides to Effective Communication

- Chase. S. *The Tyranny of Words.* New York: Harcourt Brace, 1959.
- Gordon, T. *Leadership Effectiveness Training (L.E.T.), The No-Lose Way to Release the Productive Potential of People.* New York: Wyden, 1958.
- Delbecq, Van de Ven, Gustafson. *Group Techniques for Program Planning.* Glenview, Ill.: Scott, Foresman and Co., 1975.
- Hegarty, E.J. *How to Run Better Meetings.* New York: McGraw-Hill, 1957.
- Mager. R. *Preparing Instructional Objectives.* Belmont, Calif.: Fearon-Pitman, 1975.
- Strunk and White. *The Elements of Style.* New York: Macmillan, 1962.

Guides to Systems Analysis

- Forrester, J.W. *Industrial Dynamics.* Cambridge, Mass.: MIT Press, 1961.
- Forrester, J.W. *Urban Dynamics.* Cambridge, Mass.: MIT Press, 1969.
- Hatry, H.P., Blair, L.H., Fisk, D.M., Greiner, J.M., Hall, J.R., Jr., Schaenman, P.S. *How Effective Are Your Community Services? Procedures for Monitoring the Effectiveness of Municipal Services.* Washington, D.C.: The Urban Institute and The International City Management Association, 1977. (Read especially Chapter 2, Solid Waste Collection.)
- Moroney, M.J. *Facts From Figures.* New York: Penguin, 1952 (data analysis).
- Sutherland, J. *Systems: Analysis, Administration and Architecture.* New York: Van Nos Reinhold, 1975.

Guides to Organization Management

- Drucker. P. *The Concept of the Corporation*. Scranton, Pa.: John Day, 1972.
- Drucker. P. *Managing for Results*. New York: Harper and Row, 1964.
- Koontz. H. *Appraising Managers as Managers*. New York: McGraw-Hill, 1971.
- Mackenzie. R.A. *The Time Trap*. New York: McGraw-Hill, 1975.
- Sloan. A. *My Years With General Motors*. Garden City, N. Y.: Doubleday. 1972.
- Townsend. P. *Up the Organization*. New York: Knopf, 1970.

Guides to Government Agency Management

- Drucker. P. *Management: Tasks, Practices, Responsibilities*. New York: Harper and Row, 1974. (Especially the chapter on nonprofits.)
- Malek. F.V. *Washington's Hidden Tragedy: The Failure To Make Government Work*. New York: Free Press. 1978.
- Niskanen. W.A. *Bureaucracy and Representative Government*. (Out of print)

Although this monograph has been the initial attempt to provide a practical guide to evaluability assessment, the subject has been treated by other professionals in a variety of ways. The 10 documents listed below comprise a partial listing of publications in which efforts have been made to define and analyze this many-faceted subject.

- General Accounting Office. *Finding Out How Programs Are Working: Suggestions for Congressional Oversight*. November 1977.
- Scanlon. J.W. *Evaluability Assessment: Avoiding Types III and IV Errors*. Washington. D.C.: The Urban Institute. 1977.
- Scanlon. J.W. *Federal Evaluation Policy: The Cart Before the Horse*. Paper presented at the ORSA/TIMS Conference. San Francisco. May 10. 1977
- Schmidt. E.. Horst. P.. Scanlon. J.W.. Wholey. J.S. *Serving the Federal Evaluation Market*. Washington. D.C.: The Urban Institute. 1977.
- Wholey. J.S. *Evaluation: When Is It Really Needed?* Evaluation Magazine. 2(2) 1975.
- Wholey. J.S. *A Methodology for Planning and Conducting Project Impact Evaluations in UNESCO Fields*. Washington. D.C.: The Urban Institute, 1976

- Wholey, J.S., Bell, J.B., Scanlon, J.W , Waller, J.D. *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Bureau Program I: Rulemaking—Interpreting the Law at the Federal Level.* Washington, D.C.: The Urban Institute, November 1977.

- Wholey, J.S., Bell, J.B., Scanlon, J.W., Waller, J.D. *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Bureau Program II: Administering the Law Nationwide.* Washington, D.C.: The Urban Institute, November 1977.

- Wholey, J.S., Bell, J.B., Scanlon, J.W., Waller, J.D. *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Bureau Program III: Developing and Disseminating Discretionary Technical Assistance Products.* Washington, D.C.: The Urban Institute, November 1977.

- Wholey, J.S., Bell, J.B., Scanlon, J.W., Waller, J.D. *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Approach: Method and Programs Selected.* Washington, D.C.: The Urban Institute, December 1977.

REFERENCES

1 Hatry, H.P., Blair, L.H., Fisk, D.M., Greiner, J.M., Hall, J.R., Jr., Schaenman, P.S. *How Effective Are Your Community Services? Procedures for Monitoring the Effectiveness of Municipal Services.* Washington, D.C.: The Urban Institute and The International City Management Association, 1977.

2. Horst, P. *Program Description and Evaluation Plan for the Montgomery County MCI Program.* Washington, D.C.: The Urban Institute, 1978 (internal working paper).

3. Horst, P., Scanlon, J.W., Schmidt, R.E., Wholey, J.S. *Evaluation Planning at the National Institute of Mental Health: A Case History.* Washington, D.C.: The Urban Institute, 1974.

4 Horst, P., Nay, J.N., Scanlon, J.W., Wholey, J.S. *Program Management and the Federal Evaluator.* Public Administration Review, July/August 1974, p. 6.

5. Nay, J.N. *The Development and Use of Measurement Models: A Short Feasibility Study Based Upon Four Completed Phase I Reports, Annex D to the National Institute's Information Machine.* Washington, D.C.: The Urban Institute, 1977.

6 Nay, J.N., Scanlon, J.W., Graham, L., Waller, J.D. *The National Institute's Information Machine: A Case Study of the National Evaluation Program.* Washington, D.C.: The Urban Institute, 1977.

7 Schmidt, R.E., Scanlon, J.W., Horst, P., White, T., Tash, W.R. *Appalachian Regional Commission Health and Child Development Program: Analysis of Evaluation Prerequisities (Vol. I).* Washington, D.C.: The Urban Institute, 1976.

8 Wholey, J.S. *Evaluation: Promise and Performance.* Washington, D.C.: The Urban Institute (in press).

9 Wholey, J.S., Bell, J.B., Scanlon, J.W., Waller, J.D. *Evaluability Assessment for the Bureau of Health Planning and Resources Development, DHEW: Decision Paper.* Washington, D.C.: The Urban Institute, December 1977