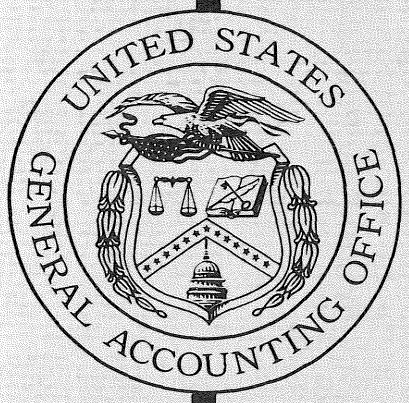


COMPREHENSIVE APPROACH FOR PLANNING AND CONDUCTING A PROGRAM RESULTS REVIEW

EXPOSURE DRAFT

UNITED STATES
GENERAL ACCOUNTING OFFICE

JUNE 6, 1978

UNITED STATES GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

DIVISION OF FINANCIAL AND
GENERAL MANAGEMENT STUDIES

JUN 6 1978

The focus of government auditing has expanded. Auditing is no longer merely concerned with the appropriateness of financial records and compliance with legal requirements. Now the auditor is also being called upon to assess the economy, efficiency, and effectiveness of program operations. These additional responsibilities are not only new, they are also challenging.

This document addresses the area of program effectiveness or more specifically, whether a program is achieving a desired level of results. Reviewing program effectiveness is by no means a clear-cut task. Unlike the relatively stable environment of financial auditing, audit organizations have tended to approach program results review with some uncertainty. The uncertainty can be categorized in terms of both "how" the review should be conducted and "what" types of performance information are needed to satisfy the review objectives.

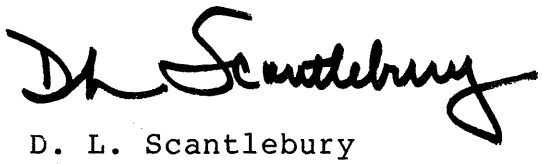
To minimize the uncertainty, both conceptual and practical guidance are provided in this guide. The conceptual guidance focuses on the salient elements of effectiveness measurement systems and their impact on the program results review. The practical guidance involves a comprehensive approach for planning and conducting reviews that is consistent with the intent of the yellow book standards. ("Standards for Audit of Governmental Organizations, Programs, Activities & Functions.") Explanatory examples are also included to further illustrate specific issues or procedures.

This document was prepared by a team from the San Francisco Regional Office with advisory assistance provided by a subcommittee of the Western Intergovernmental Audit Forum. Appreciation is extended to the numerous Federal, State, and local audit and evaluation organizations, CPA firms, and professional organizations for their comments and examples.

Like its companion document, "Audit Guidelines for Audits of Financial Operations of Federally Assisted Programs," this document is being issued as an exposure draft to all members of the National and Regional Audit Forums and other individuals in the auditing community. We encourage your use of this document and solicit your comments on how it can be improved. Please send these comments to:

D. L. Scantlebury, Director
U.S. General Accounting Office
Financial and General Management Studies
Division
441 G Street, NW.
Washington, D.C. 20548

Sincerely yours,



D. L. Scantlebury
Director

Contents

LIST OF FIGURES AND TABLES

ABBREVIATIONS

GAO	General Accounting Office
RFP	request for proposal

CHAPTER 1

INTRODUCTION

BACKGROUND

The program results review is one of many new approaches that focus on the accomplishments and potential of publically funded programs.

Although program accountability is not new, government oversight bodies are considering new ways to obtain objective program performance information. This fresh look at program accountability is due, in part, to the proliferation and increased cost of public programs, as well as the current lack of consistent and complete internal performance reporting. To improve oversight, government analysts are designing and testing more uniform and comprehensive monitoring devices. Two examples that have received sporadic attention are sunset legislation and zero-based budgeting.

Auditors have also laid the necessary groundwork to provide this new dimension of program review. A 1972 GAO publication entitled Standards for Audit of Government Organizations, Programs, Activities, & Functions" (called the yellow book) acknowledged an expanded audit role within the public sector. In addition to establishing standards to improve audit quality, the yellow book defined the objectives of auditing as reviewing

--financial operations and compliance with applicable laws and regulations

--economy and efficiency of management practices, and

--the effectiveness of programs in achieving a desired level of results.

Although this last objective focuses on the program performance data needs of government decisionmakers, many audit organizations have been reluctant to undertake audits in this new area. One reason for their reluctance stems from a general lack of understanding as to what is required and how it can be accomplished.

PURPOSE OF THIS DOCUMENT

This document has two primary purposes. First, it provides practical guidance to audit organizations with little or no program results review experience. Second, it proposes a universal approach that can be used to govern the quality and consistency of program results review assignments.

The need for practical guidance is shown by both the lack of program results review experience and the relative void in auditing literature. While much has been written in the related topic area of program evaluation, little is available that bridges the gap between program evaluation theory and its practical application in an audit environment. This document attempts to bridge that gap. It provides a practical approach together with illustrative examples for planning and conducting program results reviews.

Furthermore, this document provides government officials, who request program results reviews, with a way to enhance the consistency and elevate the overall quality of future review assignments. The lack of a universally recognized approach for conducting such reviews has precipitated numerous strategies of varying quality to satisfy the general review objectives. The resulting inconsistency tends to promote a skepticism among audit users, which in turn can hamper the credibility of even the most reliable reviews findings. When auditors use the framework proposed in this document, audit requestors can be at least partially assured that future assignments will be planned and conducted consistent with the standards contained in the gold book

GUIDELINE FORMAT

Program results reviews are conducted in an uncertain environment. The numerous variables that cause this uncertainty preclude the design of an entirely prescriptive or step-by-step guideline format. Consequently, the guidance provided in this document is more descriptive than prescriptive. That is, it describes the general process that governs the conduct of a program results review. Some review activities, nonetheless, lend themselves to more prescriptive guidance. In these instances, specific procedures have been developed to assist the reviewer.

The following chapters provide a conceptual as well as a practical framework. Chapters 2 and 3 highlight the nature of the program results review process and the importance of effectiveness measurement systems. Definitions of related terms and concepts are included in these two chapters. The more practical aspects of planning and conducting a program results review are in chapters 4 through 9. Each of these chapters is devoted to a major review activity.

In addition, a comprehensive flowchart that shows the interrelationship of all the program results review activities is in appendix A.

CHAPTER 2

WHAT IS A PROGRAM RESULTS REVIEW?

A program results review determines whether intended results or benefits are being achieved and, if appropriate, identifies areas for improvement. Such assignments, however, can be better understood by clarifying the

--basic terminology,

--specific review objectives, and

--general process for planning and conducting program results reviews.

DEFINITION OF TERMS

To provide a common understanding, the terms program, program result, program effectiveness, and program results review are defined as follows.

Program

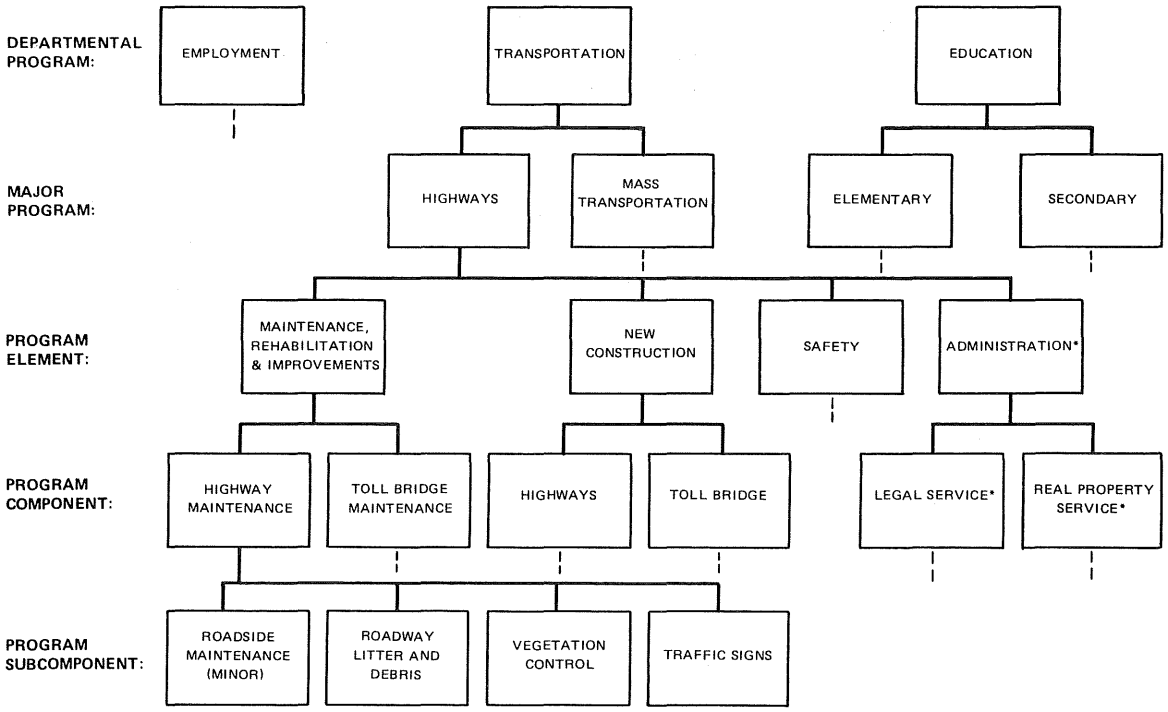
A program comprises specific activities that attempt to accomplish one or more objectives.

The term "program" is not confined to any particular level of government. As Figure 1 shows, several levels of program activity are possible, ranging from high-level departmental programs down to specific program subcomponents. Higher level programs are merely an aggregation of lower level program activities.

Furthermore, the definition does not limit the term "program" to only those activities that result in a public benefit or service. Other activities that guide and support public benefit-oriented programs also have specific objectives and are equally subject to review. Examples of support-type programs include administrative service, personnel training, equipment maintenance, and legal service.

FIGURE 1

EXAMPLE OF FIVE LEVELS OF PROGRAM ACTIVITY



***These program elements and components represent support type services that are also subject to program results review.**

Program result

A program result is the manifestation of a program objective; or more simply, it is the desired change or accomplishment that occurs as a result of a program.

By definition, a program is intended to accomplish some objective, thereby producing a desired change or program result. The following examples show this interrelationship:

Program	Objective	Result
Job Training Program	To increase the employability of disadvantaged youths.	The change in employment status within the program's target group.
Vehicular Inspection Program	To deter the operation of unsafe vehicles.	The change in traffic accidents resulting from preventable mechanical failures.

Program effectiveness

Effectiveness is a measure of the extent to which a program achieves a desired level of program results.

Effectiveness involves a comparison between what a program actually accomplishes and what it was intended to accomplish. Since precise measures of expected accomplishments are often unavailable, acceptable performance standards generally need to be developed. (Performance standards are discussed in greater detail in ch. 3.)

The following example illustrates the concept of effectiveness and its relationship to the performance standard used:

One objective for all fire departments is to contain fires to the building in which the fire originates. During a program results review, a small town's professional fire department was found to have a containment rate of 80 percent. Since the fire department's charter did not cite a specific measurable goal, an acceptable performance standard had to be developed. To determine what containment rate may be appropriate, inquiries of other comparable fire departments were made. They reported containment rates between 85 and 90 percent.

As a result, the local fire department's containment program was found to be slightly less effective than similar fire containment programs in comparable jurisdictions.

Program results review

A program results review is a process or approach by which qualified individuals can determine the level of program effectiveness and, if necessary, identify areas for improved program performance.

A program results review extends beyond traditional audit theory into the realm of activities commonly known as evaluation and analysis. Program results review activities are neither constrained to the conventional audit of information and control systems nor as pervasive as the wide range of activities associated with evaluation and analysis. 1/ Since the terms audit, evaluation, and analysis carry implications that do not necessarily apply to the program results review process, the term "review" is used predominantly in this document.

This distinction, however, does not negate the required compliance with the yellow book's standards. As long as these standards are maintained, a review team can be composed of individuals from either the audit or evaluation communities.

REVIEW OBJECTIVES

A program results review incorporates three objectives that together satisfy the yellow book's requirements and intentions. These objectives require:

--Assessing the adequacy of management's system for measuring effectiveness.

--Determining whether a program satisfactorily achieves a desired level of program results.

--Identifying causes that inhibit satisfactory performance.

1/A detailed discussion of evaluation and analysis activities is contained in the GAO document, "Evaluation and Analysis to Support Decisionmaking" (PAD-76-9), September 1976.

These objectives are not mutually exclusive. For example, the determination of a satisfactory level of program results precludes the need to identify causes that inhibit effectiveness.

Nonetheless, the combined or comprehensive nature of these objectives recognizes various management and oversight needs. For example, a comprehensive program results review not only substantiates a program's current level of effectiveness but also enhances or endorses the program's self-measurement capabilities. In addition, a comprehensive review attempts to establish a causal relationship between level of effectiveness and the factors that inhibit increased performance. The development of this causal relationship reduces the uncertainty that too frequently accompanies management attempts to improve performance.

Despite the advantages of this comprehensive approach, requests for program results reviews may occasionally limit the scope of a review assignment to only one or two of these objectives. Review agencies must be alert for such limitations. Circumstances under which limited reviews may occur are discussed in chapter 4.

PROGRAM RESULTS REVIEW PROCESS

The review objectives determine both the planning and conduct of a review assignment. To satisfy these objectives, this document proposes a systematic process that consists of six major activities. The flow, interrelationship, and purpose of these activities are shown in Figure 2.

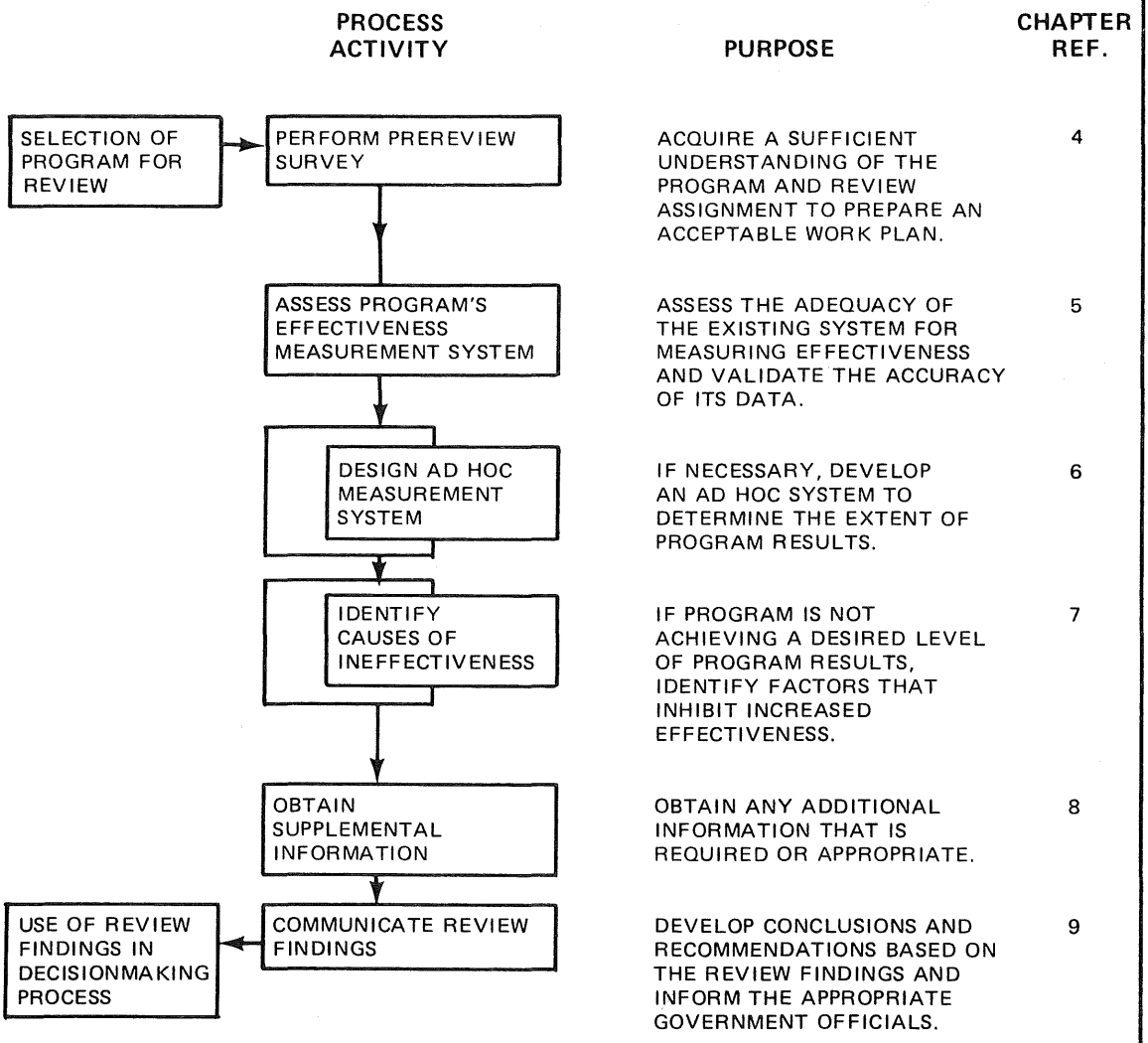
This process is not definitive. Although it reflects the most common flow of activities and is easy to understand, some adaptation may be appropriate during a review assignment. For instance, information needed to determine the level of effectiveness and to identify possible causes of ineffectiveness can often be collected concurrently. Furthermore, the supplemental information can often be obtained earlier than shown in Figure 2.

The process format in Figure 2 does not mean to suggest that all review activities are the same. The time and staff needed to accomplish each activity will vary based on a number of factors. For example, prior program exposure will affect the amount of time needed during planning and familiarization. Likewise, a reliable effectiveness measurement system will preclude the need to design an ad hoc system.

Each process activity is discussed in a separate guideline chapter as indicated in Figure 2.

FIGURE 2

PROGRAM RESULTS REVIEW PROCESS



CHAPTER 3

EFFECTIVENESS MEASUREMENT SYSTEMS AND THEIR IMPACT ON PROGRAM RESULTS REVIEWS

To better understand how program results reviews are conducted, the review team should possess a working knowledge of effectiveness measurement systems.

Accepted management practices requires developing and using a system to measure effectiveness. Such systems basically involve charting progress toward program objectives and comparing actual achievement against what was intended. The data from these systems cannot only provide managers with feedback on the adequacy of current program operations but also highlight opportunities to improve future performance.

The use of systems to measure effectiveness is commonplace in the private sector. When profit is the primary objective, financial measurement systems can be relatively easily designed. Such systems involve compiling daily receipt and disbursement data, developing financial forecasts, and using variance analysis to compare actual profits with expected profits.

Although public sector programs do not have the convenience of the single and easily measured profit objective, systems to measure public program effectiveness can usually be designed. This chapter discusses the general characteristics and design of effectiveness measurement systems, as well as the impact these systems have on the planning and conduct of a program results review.

EFFECTIVENESS MEASUREMENT SYSTEMS

Like all systems, effectiveness measurement systems have a structure and a process. The structure describes what the system is and the process describes what the system does. Effectiveness measurement systems consist of three structural components and two operational or process activities. These five elements are shown in Table 1.

The following subsections explain the purpose of each of these five elements and the appropriate characteristics that they should possess. Specific procedures for assessing the adequacy of these elements are in chapter 5.

Table 1

ELEMENTS OF AN EFFECTIVENESS MEASUREMENT SYSTEM

I.	STRUCTURAL COMPONENTS:		
	A. Performance Indicators	-	quantifiable expressions of program objectives
	B. Data Source	-	base from which information about performance indicators can be obtained
	C. Performance Standards	-	desired level of achievement for a performance indicator
II.	PROCESS ACTIVITIES:		
	A. Data Collection Process	-	collecting performance indicator data from the data source
	B. Comparison Process	-	comparing the actual status of a performance indicator with the appropriate performance standard to determine extent of program effectiveness

Performance indicators

Performance indicators are measurable expressions of a program objective.

In rare instances a program objective may be its own performance indicator. This occurs when a program objective is explicitly stated to permit direct measurement. For example, an explicit objective established for a job placement program would be

> "to place eligible applicants (as defined by the legislation) in substantially full-time (not less than 30 hours) permanent (minimum of 6 months) jobs with a salary of not less than $500.00 per month."

Generally, program objectives are not expressed in this detail, and measurable surrogates or performance indicators need to be developed. Often multiple performance indicators are needed to capture the intent of a single program objective. For instance, a highway construction program may have as one of its mandated objectives "the improvement of highway safety." Although safety is not directly measurable, several measurable surrogates exist that reflect different aspects of the magnitude and seriousness of unsafe conditions. For example, frequency of accidents, number of fatalities, number of injuries, and dollar value of property damage could be used as multiple performance indicators to provide a reasonable approximation for the otherwise unmeasurable safety objective.

Performance indicators must always reflect the desired societal change or result intended by the program objectives. Generally, the measurement of work or program outputs is not an acceptable substitute for a program objective. For example, a reduction in the number of traffic citations does not necessarily suggest a reduction in unsafe driving. Similarly, an increase in the number of courses available through a vocational training program does not imply an increase in the employment of program participants.

When performance indicators are used as surrogates for program objectives they should possess the following qualities:

Validity - The indicators reasonably represent the program objective.

Sufficiency - The number of indicators used adequately reflects the intent of the program objective.

Data source

The data source contains or provides information about the status of a performance indicator.

Common data sources are

--program files, records, studies, etc.;

--program personnel;

--past or present program beneficiaries;

--other persons having direct program knowledge;

--statistical gathering agencies, such as the Bureau of the Census or Bureau of Labor Statistics.

For example, information on the number of highway accidents, injuries, and fatalities may be in police accident investigation reports, while property damage statistics may be available through an automobile insurance association.

Quality characteristics that apply to data sources include:

Reliability	-	The data is accurate, consistent, and dependable.
Bias free	-	The data is fair and impartial.

Performance standard

A performance standard represents the desired level of achievement for a performance indicator.

Selecting performance standards is the most elusive aspect in the design of an effectiveness measurement system. Although every program is expected to achieve some satisfactory level of performance, the legal authorities that establish programs seldom designate what level or standard of performance is acceptable. Instead, program objectives are often stated in such vague terms as "providing" or "increasing" a particular service or benefit. As a result, a whole spectrum of possible performance standards may exist.

The following example illustrates this dilemma:

A training program for disadvantaged youths is in its third year of operation. The program objective is to successfully place program participants into the job market. (Success would have to be defined, e.g., 1-month, 1-year, training-related, full-time; for this example, however, the performance indicator and standards are assumed to have comparable success definitions.) The program results review found the program to have successfully placed 65 percent of its second-year participants. The following possible standards or placement rates are available:

		Percent
a.	Ideal	100
b.	National employment rate	92
c.	Long-term goal established by program management	80
d.	Norm for comparable training and placement programs	70
e.	Placement rate for first-year participants	55
f.	Placement rate for untrained control group	40

Given these choices, the program could be determined to be either highly effective or highly ineffective depending on the performance standard used. Thus, it is necessary to determine which performance standard is the most appropriate. Generally, the following guidelines should be considered when attempting to select an appropriate performance standard:

--Satisfies the legislative intent.

--Represents or approaches the outer limit of what the program is expected to accomplish.

--Is reasonable or practically attainable.

--Is compatible with the program objective and performance indicator.

These guidelines enable the designer of an effectiveness measurement system to narrow the range of possible alternatives. From the previous example, both the ideal and national employment rates are most likely unreasonable or not practically attainable in the forseeable future. Similarly, both the placement rate for the untrained control group and last year's placement rate are not the most appropriate performance standards since they do not represent the outer limits of what the program is expected to accomplish. This information, nonetheless, is important for it shows that the program is producing a positive societal change and that the extent of this change is steadily increasing as the program matures. (These concepts will be discussed again in both chs. 5 and 7.)

Either of the remaining placements rates may be acceptable. The norm or average placement rate for the other training and placement programs may be appropriate if they are indeed comparable. However, since the 70 percent rate is a norm, some comparable programs may achieve more than

70 percent while others achieve less. If this is the case, management's long-term goal of 80 percent may be practically attainable, and thus even more representative of the outer limit of what the program is expected to accomplish.

Data collection process

Data collection is the process of compiling performance indicator data to show the actual status or achievement level for a specific time period.

Such data can be collected by one or some combination of the following methods:

--Interviews or questionnaires directed toward individuals with sufficient program knowledge to respond, e.g., program employees, program clients, or knowledgeable third parties.

--Observation or inspection of program operations or practices.

--Review of pertinent records, files, reports, or statistical data.

For example, job placement information could be solicited directly from former program participants, accident information could be extracted from investigative reports, and recreational usage data could be obtained through observation.

The quality characteristics that apply to data sources similarly apply to the data collection process. In both cases, people are providing, collecting, or otherwise manipulating data with varying degrees of accuracy, reliability, and bias. For example, both the program participants responding to a questionnaire and the analysts collecting and synthesizing their responses may be affected by subjective influences that can distort true program performance. An effectiveness measurement system should incorporate sufficient controls to minimize any possible distortion in data accuracy or reliability.

Comparison process

The final element in an effectiveness measurement system is the comparison process. This process merely involves comparing the actual level or status of a performance indicator with the appropriate performance standard. The

assurance of an accurate computation is the only quality requirement that applies to this comparison activity.

If the other system elements satisfy their quality conditions, this process finalizes the determination of program effectiveness, as defined in chapter 2. When the actual status of the performance indicator equals or exceeds the performance standard, the program is considered effective in achieving its program objective. Conversely, a program's level of effectiveness is diminished to the extent that the actual status of a performance indicator falls short of its performance standard.

Concluding remarks

The previous discussion establishes the appropriate structure, processes, and quality characteristics of an effectiveness measurement system. While the performance findings generated by such systems are critical to program managers, additional information is needed to improve future program performance. This requires identifying the causes that hinder satisfactory performance. By identifying these causes, program managers can redirect program resources or modify operating strategies to eliminate or reduce the influences of these causes. If these causes are too significant and cannot be eliminated, program termination may be appropriate.

Occasionally a mechanism for identifying causes can be built into an effectiveness measurement system. In these instances, causal information is generally colocated with performance indicator data, and the two pieces of data can be collected concurrently. For example, highway accident investigation reports contain data on both number of accidents and the reasons they happened.

In other instances, causal information can neither be obtained nor anticipated during the data collection process. A discussion of how causes can be identified, after a determination of ineffective performance is made, is in chapter 6.

IMPACT OF EFFECTIVENESS MEASUREMENT SYSTEMS ON THE CONDUCT OF PROGRAM RESULTS REVIEWS

The scope of a program results review is determined largely by the adequacy of management's effectiveness measurement system.

Managers are generally free to determine effectiveness by whatever means they believe is appropriate. Systems may range from those that are formal and well designed and reflect the quality characteristics discussed previously to those that are informal and rely heavily on management intuition. The resulting variation creates an uncertain environment for the program results review.

Unlike program results reviews, financial audits are planned and conducted in a more stable environment. Similar to the assessment of management's effectiveness measurement system, a financial audit concentrates on the adequacy of management's financial accounting system and the reliability of that system's data. A well-defined body of principles and standards, however, governs the design and use of financial accounting systems. The existence of such highly structured and universally accepted systems for recording and developing financial data reduces the level of uncertainty confronting the financial audit team. Thus, audit activities such as assessing the adequacy of financial controls can be relatively easily planned and conducted. Furthermore, if financial accounting system problems are observed, recommendations to correct these problems satisfy the audit responsibility.

Unfortunately, a well-defined and universally accepted body of principles and standards does not similarly govern the design and use of effectiveness measurement systems. The variation in designs and related uncertainty places a greater burden on the program results review. For example, considerable prereview planning is required to minimize or prepare for this uncertainty. In addition, if a measurement system is found to be inadequate, the review team must not only propose recommendations to correct the deficiency but also expand the scope of the review assignment. This expansion may entail either modifying management's system for measuring effectiveness or developing an ad hoc system to generate the required performance data.

As long as this uncertain environment exists, the planning and conducting of a program results review will have to consider the possible variations in effectiveness measurement systems. Three general cases are presented in Table 2 to illustrate the impact of effectiveness measurement systems on the scope of a program results review.

Concluding remark

Although a program results review may require the independent development of performance data, the design and use of an ad hoc effectiveness measurement system during a review neither usurps nor supplants management's responsibility and prerogative to develop a permanent effectiveness measurement system.

Table 2

IMPACT OF EFFECTIVENESS MEASUREMENT SYSTEMS ON THE SCOPE OF A PROGRAM RESULTS REVIEW

Quality of the effectiveness measurement system	Scope of the program results review
1. (Ideal) Effectiveness measurement systems are governed by a well defined and universally accepted body of principles and standards.	1. The review involves assessing the adequacy of the system and the reliability of its data. System deficiencies would be reported to and corrected by program management.
2. Effectiveness measurement systems vary considerably in terms of quality and completeness.	2. If the assessment of a system establishes a minor deficiency, the review staff may have to modify or supplement the existing system in order to develop a reliable determination of effectiveness.
3. Effectiveness measurement systems are nonexistent or totally inadequate.	3. The review includes the design and use of an ad hoc effectiveness measurement system to satisfy the review objective.

CHAPTER 4

PREREVIEW FAMILIARIZATION AND PLANNING ACTIVITIES

Although program results reviews are conducted in an uncertain environment, the review team can generally reduce or at least prepare for this uncertainty before accepting or beginning a review assignment. This involves acquiring an early awareness of the factors that affect assignment performance. This chapter addresses these factors in conjunction with the following prereview planning and familiarization activities:

--Acquire a complete understanding of the assignment objectives and requirements.

--Obtain requisite program familiarity.

--Determine whether to pursue the review assignment.

--Select an appropriate review strategy.

--Prepare a work plan or bid proposal.

ASSIGNMENT CLARIFICATION

Before initiating a program results review, the review team must clearly understand the specific objectives and unique requirements of the review assignment. This information is normally provided in the document authorizing the review. For review agencies with the authority to self-initiate assignments, this information should be provided in an initial justification document. When assignments are not internally initiated, this information should be contained in the review request or request for proposal (RFP) prepared by the agency or body authorizing the review.

Occasionally, the necessary information is not adequately explained in the assignment request or justification document. In these instances, the review team should solicit further clarification from those officials authorizing the review. Regardless of the information source, the staff assigned to prepare the work plan and conduct the review must know exactly what is expected.

Key issues to clarify during this assignment familiarization activity include (1) specific assignment objectives,

(2) basic review and reporting requirements, (3) identification of program objectives, and (4) resolution of potential problems unique to program results reviews.

Assignment objective

A review assignment may have as its objective a comprehensive program results review, a limited program results review, or a program results review in conjunction with other yellow book audit objectives.

As discussed in chapter 2, the comprehensive approach for conducting a program results review has three subobjectives. These are (1) assessing the adequacy of management's system for measuring program effectiveness, (2) determining whether the program is achieving its intended results, and (3) identifying the factors that inhibit satisfactory performance (if the program is not operating at a desirable level of effectiveness).

Occasionally, a review assignment may require only a limited program results review that concentrates on one or two of the subobjectives discussed above. The following examples illustrate special situations that may prompt limited review requests:

Assess adequacy of
effectiveness measurement system

A city council confronted with questionable performance reports from a municipal program may desire an independent assessment of that program's system for measuring effectiveness. The review findings can either substantiate the adequacy of the existing system or recommend improvements to ensure that future reports provide more reliable performance information for city council oversight needs.

Determine whether the program
is achieving the desired results

A county government has sponsored two uniquely designed experimental programs with the same societal and program objectives. At the end of a trial period, an independent review is performed that focuses exclusively on the success of the respective programs. The review findings would provide county officials with useful information to aid in their choice of which program to continue to sponsor.

Identify factors that inhibit effectiveness

A State program is universally known to be ineffective in providing the desired societal change. Program personnel, oversight officials, the news media, and the citizenry are all aware that the program is not performing satisfactorily. Such a situation could prompt a limited program results review that concentrates solely on identifying causes of program ineffectiveness.

Although limited reviews are occasionally requested, such reviews are generally not as meaningful as comprehensive program results reviews. If a review agency receives a request for a limited review, it should assess the desirability of a more comprehensive approach and, if appropriate, discuss the alternatives with the officials authorizing the review.

In addition to either the comprehensive or limited review discussed above, an assignment request may require one or more of the other audit objectives defined in the yellow book. If financial, compliance, or economy and efficiency audit work is also requested, the review team will have to organize their work plan accordingly. 1/

Basic review and reporting requirements

Certain review and reporting requirements generally apply to all review assignments regardless of the specific objectives. For convenience, they are listed below. If they are not adequately explained in the authorizing document, they should be clarified with the appropriate government official.

--Elementary background information identifying the authority and scope of the program to be reviewed.

--Time requirements or deadlines for work plan or bid proposal as well as interim and final reports.

1/Guidelines for each audit or review element defined in the yellow book will be issued by us in conjunction with either the National or Regional Intergovernmental Audit Forums. To date, only the "Audit Guidelines for Audits of Financial Operations of Federally Assisted Programs" has been published.

--Cost or resource limitations.

--Special work plan, bid proposal, or report format requirements.

--Identification of liaison for additional information.

Program objectives

The most critical ingredient in a program results review is a clear statement of the program objectives for which the level of achievement is to be measured. The officials authorizing the review should identify these objectives in the assignment request or justification document. In addition, they should either provide or reference the enabling documents in which the specific objectives were originally established. If the program objectives are not clearly ascertainable from these documents, the review team must solicit further clarification before proceeding with the review.

Potential problems

The lack of an objective set of standards or principles to govern effectiveness measurement systems creates the potential for two unique problems that may affect the conduct of a program results review assignment. Although these problems may never surface, they should be anticipated and methods for resolving the potential problems should be clarified before beginning the review. These potential problems involve:

--Irreconciliable differences between the review staff and program management over the appropriateness of the system used to measure effectiveness.

--Contingent work responsibilities that are not readily identifiable before preparing the work plan or bid proposal.

The subjective nature of measuring program effectiveness may lead to irreconciliable differences between the review staff and program management. The appropriateness of specific performance indicators, data sources, and performance standards is determined primarily by their relationship to the program objectives. Unfortunately, this relationship is not always precise. In the absence of such precision, reasonable approximations must be considered. Reasonableness, however,

is subject to judgment, which in turn can lead to differences of opinion, as the following example illustrates:

> A Community Legal Defense program in a large metropolitan city has the objective "to provide criminal legal representation to indigent residents." Since the program did not have a system for measuring effectiveness, the review staff had to design an ad hoc system. The review staff decided to use the accomplishments of the city's Public Defender's Office as the performance standard. Program management opposed this performance standard, maintaining that it was not sufficiently comparable. The arrangement between the review agency and the review requestor, nonetheless, permitted the review agency to assert and defend whatever position they believed appropriate. Because the review agency was unable to develop a more suitable performance standard, they continued to use the Public Defender's Office. Program management was later given the opportunity to formally rebut any assumption or methodologies used during the review to determine the level of program effectiveness.

When judgments differ and agreement cannot be reached, some means to resolve the resulting conflict must prevail. Following are three solutions for handling potential conflict that should be considered and clarified with the review requestor during this prereview period:

- Conflict is appealed to a higher oversight or moderating authority for resolution if the need arises.

- Review agency is granted the authority to assert and defend whatever position it believes appropriate.

- No further work is done, and an informational report outlining the source of the conflict is prepared.

A second problem that often affects a program results review is the inability of the review agency to precisely forecast the amount of time and resources necessary to satisfactorily complete a review assignment. This problem specifically concerns the unexpected need for additional

work should some aspect of management's effectiveness measurement system prove to be inadequate. An example of this type of dilemma follows:

> An accounting firm responding to a request for proposal to determine the results of a transit system's on-time performance program is informed that program personnel already collect and evaluate on-time performance data. Under this premise, the firm prepares a bid proposal that does not include an expenditure of time or resources to develop or collect new performance data. After the proposal is accepted and the review is initiated, the review team finds that due to poorly written procedures, the performance data collected by program personnel is both inaccurate and inconsistent; it is, therefore, inadequate for performance measurement purposes. The review agency is now confronted with the need to develop an alternative approach for determining program effectiveness. This, however, will involve considerable additional work not provided for in the original proposal.

Because unanticipated work may be required, the review agency together with the review requestor should develop a mutually acceptable contingency work plan. Since little experience has been gained in this area, the initial arrangements will of necessity be innovative or experimental. Two possible suggestions are:

--Work plans or bid proposals that stipulate in advance the amount and cost of additional work assuming different contingency situations.

--Two-stage work plans or bid proposals that provide for a renegotiation following the assessment of management's effectiveness measurement system.

PROGRAM FAMILIARIZATION

The second major step in the planning process involves quickly acquiring a background knowledge of both general program operations and specific methods for measuring effectiveness. This activity is essentially a preliminary survey. 1/ The extent of such preliminary work depends on review agency's prior program knowledge and the availability of program information.

1/The survey is discussed in detail in Audit Standards Supplement Number 11.

The type of information needed is similar for all programs and involves three general areas of inquiry, specifically

--verification of program objectives,

--identification of effectiveness measurement systems, and

--familiarization with program operations or activities.

The review agency should verify the program objectives identified during the assignment clarification process with program management. Since program objectives are often vague, differences of opinion or interpretation may arise. If any variances or misunderstandings are noted, they will have to be resolved. This may involve recontacting the officials who authorized the review assignment. The following example demonstrates how management's view of program objectives can differ from those objectives originally established for a program:

> The National School Lunch Program, initiated in 1946, had two objectives: (1) to provide nutritious meals to the Nation's children and (2) to supplement farm income by increasing food demand. Program management stated, however, that a change in national priorities had caused them to no longer pursue the second program objective. Although the legal authority under which the program was established had never changed, the review agency analyzed program management's position. Since the problem of agricultural surplus had in recent years been replaced by the opposite problem of agricultural shortage, the review agency suggested that the Congress reconsider the appropriateness of the program objective. Since all parties concurred with the first objective (nutritious meals), the review agency continued to examine the program's effectiveness relative to that objective.

Program familiarization also involves the identification of the program's effectiveness measurement system. It is the appropriateness of this system that will determine the review strategy and, hence, the time and resources needed to conduct the review. To acquire the requisite knowledge, the following questions should be asked during this stage:

--Does program management measure program effectiveness?

--Are all program objectives subject to effectiveness measurement?

--What performance indicators, information sources, and performance standards are used to measure effectiveness?

--How is performance data collected?

--What controls are used to ensure data accuracy?

--What information sources are available if an ad hoc method for measuring program effectiveness needs to be designed?

The following example illustrates a measurement system inadequacy that could be identified during this preliminary survey activity:

> An airmail improvement program was established to increase revenues by improving airmail service. The service objective was "next day delivery for 95 percent of zip-coded airmail deposited in specially marked containers in major metropolitan areas." To measure effectiveness, the post office mailed test letters between selected cities. A preliminary observation of the data collection process revealed that the test letters were readily identifiable and, thus, subject to priority handling. Based on this suspected deficiency, the review agency was able to (correctly) anticipate the need to collect new performance data by mailing unidentifiable test letters.

A working knowledge of program operations or activities will also enhance assignment planning. This knowledge can be gained through prior program exposure, review of publically available program literature (for example legislation), interviews with key program personnel, observation of program activities, and review of program budgets, organizational charts, policy statements, procedural manuals, and performance reports. The advantages of prereview program familiarization are illustrated in the following example:

> A bilingual education program was charged with three objectives: (1) identify effective bilingual education approaches, (2) adequately train bilingual

education teachers, and (3) develop suitable instructional materials. During the prereview familiarization process, the review staff discovered that no plans had been developed to carry out, evaluate, or monitor the program. Based on this initial exposure to the program, the review team recognized a high potential for ineffective performance. This lack of adequate program planning was later substantiated to be the main factor inhibiting a satisfactory level of program effectiveness.

DECISION TO CONTINUE

Having clarified the assignment request and acquired preliminary knowledge of the program, the review agency should assess the desirability of pursuing the review. Continuation or termination decisions hinge on both the amenability of the specific program to review, the capability of the review agency itself, and the relationship between the review agency and the review requestor.

A review agency may decide not to pursue a particular assignment based on the program environment. Table 3 outlines specific situations that may invalidate the need or desirability of a comprehensive program results review.

A review agency must also consider its own ability to perform the program results review. The agency must either have staff with the necessary experience or education or be willing to hire outside consultants. If a review agency determines an assignment to be too complex or technical, it may decide not to pursue that particular request.

The relationship between the review agency and the review requestor may determine what action is appropriate if one of the preceding problems is encountered. Generally, private firms responding to an RFP in a competitive environment are not able to negotiate changes in the assignment scope or objectives. They are, therefore, faced with either limiting the scope of their bid proposal to reflect what they believe is possible or appropriate, or not responding to the RFP. Public audit or review agencies, on the other hand, must justify their reasons for not wanting to pursue an assignment request. Based on their justification, the review requestor may (1) cancel the assignment, (2) modify the assignment request, (3) authorize additional staff or consultants for the review agency, (4) submit an RFP to a private audit or review agency, or (5) permit the public agency to pursue the assignment as requested, recognizing the undesirable condition and the potential qualification of findings that may arise.

Table 3

SITUATIONS THAT DIMINISH THE DESIRABILITY OF A COMPREHENSIVE PROGRAM RESULTS REVIEW

Condition	Examples
1. Program is too immature to be reviewed.	1. A 3-year training program is only in its second year of operation.
2. Past performance data does not reflect current operations.	2. Having recently recognized its own weaknesses, program management implemented major new changes in operating policies and practices. (Note: a limited scope review concentrating solely on the identification of causes of ineffectiveness and validation of management's changes may possibly be appropriate.)
3. Program objectives are too imprecise or controversial.	3. A program objective, such as "to improve the quality of life," is too general and hence precludes universally acceptable indicators.
4. Lack of measureable performance data.	4. Pure research and development programs with the objective of advancing the state of the art cannot generally be measured.
5. Inability to demonstrate causal relationship between specific program and societal change.	5. A general tax rebate program may have as one of its objectives a decrease in the unemployment rate. Such a change would be impossible to directly attribute to this program on an aggregate basis since variables such as other monetary or fiscal strategies, the demand for new employees, the supply of trained workers, and various social programs with employment objectives would also affect changes in national employment statistics.

SELECTING A REVIEW STRATEGY

Having decided to continue, the review agency can proceed to the next major planning activity: the selection of an appropriate review strategy. The strategy which is selected will depend on the review agency's initial assessment of the program's effectiveness measurement system.

In many respects, the selection of a review strategy parallels the conceptual relationship between effectiveness measurement systems and program results reviews outlined in Table 2.

Because the information obtained prior to commencing a review may be limited or imperfect, the strategy selected is not intended to imply a rigid review structure. Instead, it provides a tentative direction or approach from which initial time and resource requirements can be developed for inclusion in the bid proposal or work plan.

The following sections highlight three separate review strategies as well as the possible need for multiple strategies.

Assess management's effectiveness measurement system

This strategy concentrates on an assessment of management's effectiveness measurement system. It is an appropriate strategy when information obtained during the preliminary survey suggests that management uses a well-designed system to measure the achievement of all program objectives. Although this may appear to be the best strategy based on the preliminary inquiry, the system may later be found to be deficient and, thus, necessitate additional review work. As discussed on page 22, contingencies for such additional work should be resolved before preparing the work plan or bid proposal.

This strategy is also appropriate when the review assignment is limited to only an assessment of management's effectiveness measurement system. Under this strategy, the detection of deficiencies would not require additional work. If, however, possible causes of ineffectiveness are identified during the assessment process, they should be communicated to the review requestor.

Assess and modify management's effectiveness measurement system

This strategy recognizes or anticipates minor inadequacies in the existing measurement system. Based on the preliminary survey, the review agency believes that the existing system will have to be modified to either collect additional performance data or measure effectiveness using a more appropriate performance standard. This additional work will have to be reflected in the work plan or bid proposal.

The possibility for contingent work requirements may still exist and should be appropriately resolved before beginning the review.

Develop an ad hoc measurement system

If the preliminary investigation concludes that management either does not have a system to measure program effectiveness or that such a system is totally inappropriate, the review staff will have to develop an ad hoc system to collect and measure performance data. Although the ad hoc system should be designed in line with program personnel, the review agency will have to estimate the time and resources needed to measure program effectiveness and, thus, satisfy the assignment objective.

Multiple strategies

Multiple strategies consisting of two or more of the previous strategies may be appropriate for the review of a program with multiple objectives. Based on the preliminary information, management's system for measuring one objective may be well designed, while another objective may not be measured at all. Thus, the work plan or bid proposal needs to specify the strategy to be used for determining program effectiveness relative to each program objective subject to review.

DEVELOPMENT OF WORK PLAN OR BID PROPOSAL

The final prereview step involves synthesizing and writing the results of the previous familiarization and planning activities. This document may be either a work plan or bid proposal, or both, depending on the needs of the review agency and the requirements of the government body authorizing the review.

Bid proposals and work plans serve two essential communication purposes. One, they provide a permanent record of the proposed program results review activity for higher level review and approval. Second, they provide a common understanding and direction for all review participants.

To satisfy these purposes, the following information should ordinarily be included in the work plan and/or bid proposal:

--A summary of basic program information (for example, legal authority, program operations, and measurement systems used).

--An explanation of the review strategy selected and potential for contingency work.

--An elaboration of the specific procedures to be followed by the review team.

--A statement of previously agreed upon solutions to potential conflicts or contingency work requirements.

--A timetable for interim and final reporting.

--An estimate (or proposal) of review cost.

--A discussion of report format and general content.

--Any specific requirements stated in the authorizing document.

SUMMARY

To adequately prepare for a program results review assignment, the review agency should:

--Obtain a clear understanding of the assignment request, including specific objectives and other basic or special requirements.

--Acquire a preliminary familiarity of program operations and systems used to measure effectiveness.

--Assess the desirability of pursuing the review.

--Select an appropriate review strategy.

--Prepare a work plan and/or bid proposal.

CHAPTER 5

ASSESSING MANAGEMENT'S EFFECTIVENESS MEASUREMENT SYSTEM

When management uses a system to measure program effectiveness, both the adequacy of the system and the accuracy of its data must be examined. From this examination the review team can establish the reliability of management's performance reports.

This chapter discusses the process of examining management's system and data in terms of the following general activities:

--Identifying and documenting management's system for measuring effectiveness.

--Assessing the validity and sufficiency of the performance indicators.

--Assessing the accuracy of the performance indicator data.

--Assessing the appropriateness of the performance standards.

--Determining effectiveness based on management's system.

IDENTIFYING AND DOCUMENTING MANAGEMENT'S MEASUREMENT SYSTEM

The background information obtained during the planning phase should be supplemented by the full documentation of management's measurement system. In particular, the review team should trace each objective through the system to identify the corresponding performance indicators, performance standards, data sources, and data collection techniques. If a program objective is not subject to effectiveness measurement by program management, an ad hoc measurement system may need to be designed. (See ch. 5.)

To determine the adequacy of management's effectiveness measurement system, each element that comprises the system must be assessed separately. The following sections outline the major issues a review team should consider when assessing performance indicators, data sources and collection techniques, and performance standards.

ASSESSING THE PERFORMANCE INDICATORS

Performance indicators should be valid and sufficient in relation to the program objective.

Assessing indicators for validity

A valid performance indicator should reflect the intent of the program objective. Thus, it should generally not measure workload data. In addition it should minimize the measurement of changes or results not attributable to the program.

The following discussion and examples serve to clarify these requirements.

A performance indicator should generally not be a measure of workload activity

The actual work or output produced by a program can be a useful measure of the size and scope of a program. Workload data, however, is generally not an acceptable substitute for a program objective, especially if there is insufficient evidence to show that the work or output produced by a program actually manifests itself in the desired program result. On the rare occasions where a definitive link has been authoritatively established between a change in program output and a change in program results, workload data may represent valid indicators of program performance.

The following examples illustrate the inappropriate use of workload measures as performance indicators:

> The results of a school lunch program intended to improve nutritional status should not be measured by the number of meals served. Valid indicators would measure levels of nutritional status, such as the presence of certain required vitamins and minerals in the foods consumed.
>
> A program with an objective to improve the competitive position of minority-owned business firms should not be measured in terms of number of firms assisted or number of loans guaranteed. Rather, surrogate measures of competitiveness, such as improvement in net income or net worth, might be appropriate.

A performance indicator should minimize the influence of related factors

A performance indicator should be able to distinguish the results caused by a program from those attributable to other factors. If the nature of the program precludes the use of a "pure" performance indicator, the data used to measure performance should be qualified in light of the other factors that might have changed program results.

The relationship of the performance indicator to the specific program under review is illustrated in the following examples:

> A police department is typically assigned the duty of preventing crime. Although the actual number of crimes prevented cannot be measured, changes in the number of reported crimes and estimates for unreported crimes can be used as performance indicators. If the incident of crime is reduced, however, it may be inappropriate to assign the entire change in the crime rate to the efforts of the police department. Other factors, such as the installation of better street lighting, poor weather conditions, a news media campaign to alert citizens of methods to reduce the temptation of crimes like burglary, may have also influenced the prevention of crime. Although the police department objective was achieved, the relative influence of the related factors should also be considered and reported.
>
> The performance indicators for a vocational training program should measure the number of individuals gaining employment based on the vocational skills acquired. For example, an individual trained to be a carpenter who acquires a job as a bank cashier should not be counted as a program result. This information, however, can be reported as statistical data and possibly labeled as indirectly related to the program.

Assessing indicators for sufficiency

When indicators relating to a particular objective are grouped, they should provide a comprehensive measure of effectiveness. The number of indicators required to sufficiently represent a program objective is largely a matter of need; that is, additional indicators are appropriate if they provide new information that is crucial to the determination of effectiveness. The need for supplemental indicators may be identified by reviewing the program objectives, consulting authoritative sources, and studying comparable programs.

Although multiple indicators can enhance or at least corroborate an effectiveness determination, their use will often increase the costs of data collection. The review team should be sensitive to these costs and not advocate the use of additional indicators unless they materially improve the measurement of results.

The following examples demonstrate the potential inadequacy of a single performance indicator:

> A national passenger train service has an objective to provide an acceptable level of on-time performance. If the train routes have major intermediate stops, a single indicator that only measures arrival times at the final destinations would be insufficient. A train could provide sporadic performance between the intermediate stops and still arrive at its final destination on time. Therefore, multiple indicators measuring performance between various intermediate as well as final stops would be appropriate.
>
> A program designed to protect the rights of migrant farm workers should not depend solely on the number of reported and verified violations. If possible, an additional performance indicator that estimates the incidence of unreported violations should also be developed.

Procedures following the assessment of performance indicators

Based on the previous assessment, the review team can either demonstrate the validity and sufficiency of the performance indicators or identify inadequacies that must be resolved. When the indicators are found to be acceptable, the review team can proceed to the assessment of data sources and collection techniques.

If the performance indicators are deficient, program management should be informed of the nature of the deficiency and their agreement and cooperation should be solicited to modify or supplement the existing measurement system. This modification requires the inclusion of additional or more valid performance indicators along with the appropriate performance standards and mechanisms for data collection. If management agrees to make the necessary modifications, the review agency should monitor the changes to ensure they are properly implemented.

The performance indicator used to measure the effectiveness of a student loan program was the number of loans authorized during the school year. Since the program's objective was to help students remain in school, the indicator was only a measure of workload data and not useful for gaging program effectiveness. The review team brought the deficiency to management's attention along with a proposed set of acceptable indicators. After discussing the matter, management adopted the proposed indicators and agreed to collect the needed data. One of the agreed upon indicators measured the number of loan recipients completing the school year. A second indicator recorded the number of recipients who eventually graduated. To provide additional evidence, a sample of recipients were interviewed to determine how useful the loans were for helping the students stay in school.

If management does not agree to the need for different or additional indicators, the review agency will have to proceed based on the agreements reached prior to initiating the review. (See p. 22.) This may involve terminating that portion of the review affected by the deficient indicator or designing an ad hoc system to develop the needed information. The decision to terminate would require the preparation of a justification statement outlining the deficient condition and its impact on the determination of effectiveness.

ASSESSING THE DATA SOURCE AND COLLECTION TECHNIQUES

The performance indicator data contained in a data source and collected for measurement purposes should accurately reflect the status of the performance indicator.

Data used to measure effectiveness can be obtained from numerous sources. The review team should determine whether the data contained in or provided by a data source is sufficiently reliable and unbiased.

In addition, the review team should examine the integrity of the data collection process to ensure that the data collected from the data source is not distorted. This integrity can normally be established by assessing the adequacy of management's internal controls over the data collection process.

Assessing a data source for reliability

The performance indicator data used to measure effectiveness must be reliable; that is, the data must not materially

distort the determination of effectiveness. Reliability includes accuracy, consistency, and dependability. If, during the review, any of these factors are found to be missing or questionable, the reliability of the performance data as well as the determination of effectiveness is compromised.

The following examples illustrate questionable data sources.

> The number of entries in a visitor's roster was used to calculate usage at a recreational facility. An observation of the habits of visitors at the facility revealed that many individuals did not sign or even notice the visitors' roster. In addition, some families used only one entry for the entire family, while in other families each member signed individually. As a result, the usage data was inconsistent and thus unreliable.

> One performance indicator used to evaluate the effectiveness of a county vehicle maintenance program is the number of mechanical failures reported to a central data processing office by the vehicle drivers. Interviews with a random sample of drivers disclosed substantially different opinions as to what constituted a mechanical failure. For example, some drivers had erroneously reported normal maintenance items like weak brakes as mechanical failures. Thus, the data used to calculate effectiveness was inaccurate in its present form.

Assessing the data source for bias

Bias also affects the usefulness of performance data. Bias can occur unintentionally or it may be due to subjective influences that are difficult, if not impossible, to prevent. Normally, indicators that require the opinions or judgments of program personnel should be questioned. Individuals served by a program or independent third-parties are generally more reliable data sources. Regardless of the data source, the review team should be alert for the possibility of biased performance data.

> Attendants at a recreational facility were required to manually tabulate usage data on 1 day each week. Each quarter, this data was submitted to an administrative assistant who extrapolated a quarterly attendance record from the 12 or 13 daily figures. An examination of the tabulation process disclosed that either Saturday or Sunday was used as the usage tabulation day at least six times each quarter. Since

usage was significantly higher on the weekends, the performance data was unacceptably biased.

Transit systems generally have an objective to provide courteous service to the riding public. If the performance indicator is the number of discourteous incidents, a log of such incidents maintained by each bus driver would most likely be inappropriate. Instead, a survey of individuals using the system would be a more reliable data source.

Assessing the integrity of the data collection process

A review of management's internal controls will establish the integrity of the techniques used to collect performance indicator data. To assess the adequacy of the collection techniques the review team should consider the following factors:

1. The statistical sampling techniques should be suitable

In general, performance data can be collected by ongoing systems that continuously record performance data or by using sampling procedures. If sampling is used, the assessment should include a review of the procedures followed by program personnel to ensure the statistical reliability of the data. Of particular concern are the sample size and methods used to select the sample.

2. The individuals collecting the data should be properly trained

Individuals can be used to collect performance data by conducting interviews, scheduling information from files or records, or observing program phenomena. To ensure the collection of accurate and consistent data, the individuals should be properly trained and provided clear instructions. This training should specifically caution individuals against subjectively influencing the content of the data which is collected.

3. The use of mechanical or electronic devices to collect data requires proper calibration

Mechanical or electronic devices can be used to collect performance data. For example, turnstiles can be used to record attendance or usage data. Such devices should be properly maintained and serviced to ensure data accuracy.

4. Rating scales or classification schemes must be appropriate

Performance data are often segmented into categories by using rating scales or classification schemes. Typically, a set of factors or characteristics is used to rate or classify different categories of data. To ensure the collection of consistent data, the rating scales and classification schemes must provide clear distinctions between the different categories. Additionally, procedures must be established to assure that data collected from separate locations or in different time periods are treated uniformly.

A physical rehabilitation program collects performance data on individuals being rehabilitated. For measurement purposes, the individuals are classified by categories based on the level of physical impairment. To ensure the consistency of the categories, mutually exclusive characteristics were used to assign the individuals. As an additional safeguard, random samplings of the classification determinations are subject to a peer or supervisory review.

Together with the Federal Government many major metropolitan areas instituted urban rat control programs. One program objective was to reduce the incidence of conditions that encourage rat harborage. To measure program accomplishment, the potential for rat harborage was classified into the following categories:

--Abandoned automobiles and appliances.

--Lumber on the ground.

--Other large rubbish.

--Weeds and tall grass.

Program employees who inspect the urban areas and collect this type of data had been properly instructed in what constitutes each of the potential rat harborage conditions.

5. The incidence of clerical errors should be minimized

Clerical errors can become a problem when information collected from source documents is transferred to summary sheets. Typical clerical errors involve transposing digits, recording wrong numbers, and dropping terminal digits. To minimize clerical errors, management should have sufficient safeguards or verification procedures to ensure that data is properly transferred from source documents.

6. The data collection technique should be appropriate

The choice of data collection techniques can affect the reliability of the performance data. For example, in-person interviews are generally considered superior for gathering data than telephone interviews. The review team should determine the appropriateness of the data collection technique in relation to the size and importance of the program.

7. The performance data collected should be sufficiently representative

When data is obtained from selected periods, the data should represent the entire time frame under review. Problems can develop when data is drawn from periods characterized by cyclical fluctuations, major changes in program operations, or excessive influences by external factors. For example:

> A program designed to reduce air pollution had specific indicators for measuring the level of pollutants. The data used to determine the status of the indicators was collected for a 2-month period during the rainy season. This information, however, is only valid for the 2-month period or possibly for the rainy season, but not for measuring conditions throughout the year.

Verifying the accuracy of performance data

The assessment of management's internal controls will determine the extent of tests that will be required to independently verify the accuracy of performance data. When management's controls prove to be adequate, the review team can restrict the scope of its tests. If, however, the internal controls are found to be inadequate, the review team must expand the scope accordingly.

To verify the accuracy of performance data, a sample should be taken from the data collected by program personnel. To the extent possible, this sample data should be traced back to the source from which it was originally collected. Then a comparison can be drawn between the performance indicator data previously collected and the sample data collected by the review team. This comparison should demonstrate whether the performance data collected and used for effectiveness measurement purposes is sufficiently accurate. Although rational judgment can often govern this comparison, the use of statistical analysis, such as chi-square, can provide an additional degree of mathematical rigor. For example:

> A vocational training program collected performance data to measure the number of individuals gaining employment. To test the accuracy of this data, the review team selected a random sample of individuals who completed the program. These individuals were then interviewed to verify the accuracy of management's data.

In some situations the review team cannot verify historical data because the measured phenomena cannot be observed or verified after the fact. Under these circumstances, the review team can design tests to verify the adequacy of current performance data. Although the tests will not verify historical data, they should provide sufficient evidence to gage the reasonableness of management's controls over previous data collection. For example:

> The Postal Service collected data by using test letters to determine if overnight delivery standards were being achieved. After assessing management's internal controls, the review team designed a statistically valid test of sample letters. The deposit and receipt of these letters were recorded to compare current performance with the historical performance reported by the Postal Service.

Procedures following the assessment of data sources and collection techniques

If the data sources and collection techniques are acceptable, the review team can endorse the adequacy of the performance data and proceed to the assessment of the performance standards.

If, on the other hand, the performance data is unacceptable in its present form, program management should be notified. With management's agreement and cooperation, the data

sources and collection techniques can be appropriately modified to generate reliable data. If management does not agree, the review agency will have to proceed based on the conflict resolution procedures previously established. This may entail either independently collecting the performance data or terminating further work relative to the deficient data and preparing an appropriate justification document.

ASSESSING THE PERFORMANCE STANDARDS

The performance standards should appropriately reflect the intended or desired program result.

Appropriate performance standards are not always discernible. In the ideal situation, the legal authority establishing a program will prescribe precise standards to define the intended program result. More often, however, programs have only broad and nonspecific objectives with no terminal or desired level of achievement. For example, objectives stated in such vague terms as "to increase employment," "to promote competition," or "to prevent crime" do not suggest what levels of performance are expected of the program. When precise expectations are not established by a legal authority, program management must assume the responsibility for developing appropriate standards in conjunction with the design and use of an effectiveness measurement system.

Some of the problems and solutions that relate to the selection of appropriate performance standards were discussed in chapter 3. (See p. 12.) To assess whether a standard selected by program management is appropriate, the review team should consider the following issues.

1. When available, achievement levels prescribed or intended by the legal authority establishing a program should be used as the performance standard

When a legal authority establishes or implies a level of intended performance, it should be used as the performance standard. The review team should research the enabling legislation, guidelines, or directives to determine if levels of effectiveness have been prescribed or intended for the program. Such research may also include interviewing those individuals who were initially responsible for developing or establishing the program. For example:

Sections of the Federal environmental legislation outline quantitative standards to be achieved in reducing pollution. These standards are a clear statement of intended results and provide targets to compare with the actual pollution levels to determine program effectiveness.

2. A standard should be selected based on objective techniques rather than the subjective opinion of program management

A variety of objective techniques are available to select performance standards. In some instances, formal models for forecasting program performance may be appropriate. If a forecasting model is used, the assumptions upon which the model was designed should be examined to detect possible bias. In other cases, authoritative sources that are independent of program management's influence may have developed appropriate performance standards. Finally, performance standards could reflect the level of accomplishment attained by comparable programs that are known to be successful.

Regardless of the technique used, it should involve some aspect of objectivity. Standards that are developed from highly subjective data, such as the opinion of program management, should be questioned. The review team should attempt to corroborate the appropriateness of subjectively determined standards. If this cannot be accomplished, the appropriateness of the standard is greatly diminished.

The following examples demonstrate objectively determined performance standards.

To forecast the expected results of a National Guard advertising program to bolster recruitment, State analysts designed a model that reflected such inputs as preference data supplied by recent recruits, the assessibility of National Guard Centers, the type of fringe benefits, and various demographic data. The target recruitment goals generated by this model represent appropriate performance standards since the input data was objectively determined.

The quality of the water supplied by a local utility district could be evaluated in relation to standards developed by an appropriate authoritative source, such as a national health association.

The incidence of residential fire in comparable communities may be an appropriate standard for judging the effectiveness of a local residential fire prevention program.

3. The standard should represent the outer limit of what the program is expected to accomplish

Occasionally, programs will use performance standards that do not represent a desired or expected level of achievement. Such standards may instead reflect the status of the performance indicator had there not been a program (for example, use of control groups), or the performance standard may merely be the previous year's level of achievement. Such standards measure progress. They show the program accomplishments in relation to the level of results that either would have occurred in the absence of the program or had been achieved by the program previously. Although this information is important and may even reflect a trend showing improved performance, such standards do not represent the outer limit of what a program is expected to accomplish. Examples follow.

In a large municipality, a comprehensive survey revealed that only 20 percent of all the elementary and secondary school students ate a balanced morning meal. To increase this percentage, a voluntary breakfast program was introduced throughout the school system. During its first year of operation, 40 percent of the students were found to have consumed a balanced breakfast. During its third and current year, the consumption rate was 50 percent. Using these statistics, the school officials reported the program to be highly successful based on its past accomplishments and progressive improvement. Such standards, however, do not necessarily reflect the outer limit of expected program performance. If, for example, the legal authority that authorized the program expected an 80 percent consumption rate to justify the program costs, the current performance would be judged unsatisfactory in achieving the desired level of program results.

4. Multiple performance standards may be more appropriate than single standards

Although the performance standard used by program management may be objectively determined, other equally or more appropriate performance standards may exist. The review team should determine whether other standards were considered by program management and examine all reasonable

methods for developing additional or more appropriate standards. Such an examination or search may reveal other standards that suggest the current standard is (1) too ambitious, (2) not sufficently indicative of reasonable expectation, or (3) is entirely appropriate.

If multiple standards of a different magnitude are found to be equally reasonable, both should be used to determine effectiveness. For example:

> The model used to predict the results of the National Guard advertising program concluded that a 60 percent increase in recruitment could be expected. A similar recruitment campaign in a neighboring State, however, only resulted in a 40-percent increase. Both of these standards may be appropriate when determining the effectiveness of the advertising program.

Procedures following the assessment of performance standards

If the performance standards are found to be appropriate, the review can proceed to the determination of program effective phase.

As before, the identification of an inappropriate performance standard should be brought to management's attention. Depending on the acceptance of this information by program management and the authority of the review agency to resolve potential conflict, either (1) management will develop an appropriate performance standard, (2) the review agency will develop an appropriate performance standard, or (3) further work relative to the deficient performance standard will be terminated.

DETERMINING PROGRAM EFFECTIVENESS

If the elements of the effectiveness measurement system are found to be either acceptable or subject to only minor modification, the review team can determine the level of program effectiveness using management's system. If, on the other hand, the existing effectiveness measurement system is too inadequate to permit modification, the review team will have to design an ad hoc measurement approach (see ch. 6), before a determination of program effectiveness can be made.

When an effectiveness measurement system is judged to be adequate and the process of comparing actual achievement

against desired is accurately computed, the review team can validate the reliability of the effectiveness findings in the program's performance reports. If minor modification to the system is required, the revised data will have to be used to determine effectiveness.

Procedures following the determination of effectiveness

In those instances when the program is operating below expectations, the review team should identify the factors that inhibit increased performance. (See ch. 7.) If, instead, the program results data is at or above the desired performance level, the review team should conclude the review after obtaining whatever supplemental information is necessary (see ch. 8) and preparing the final report (see ch. 9).

CHAPTER 6

USING AN AD HOC SYSTEM TO MEASURE EFFECTIVENESS

When management does not have an acceptable effectiveness measurement system, an ad hoc system may need to be developed to determine whether a program is achieving a desired level of program results. The decision to use an ad hoc system is made either following the planning phase (see ch. 4) or after assessing the adequacy of the existing effectivness measurement system (see ch. 5).

WHAT IS AN AD HOC SYSTEM?

An ad hoc system is used, in lieu of a permanent effectiveness measurement system, to satisfy the program results review objective. Generally, the measurements obtained from an ad hoc system are not as refined or comprehensive as would be possible from a permanently established and well-designed effectiveness measurement system. Nonetheless, an ad hoc system provides the users of the program results review findings with more reliable effectiveness data than they had previously been provided and shifts the burden for fine-tuning the performance data back to program management.

Regardless of the findings of the ad hoc system, program management should design and implement an acceptable effectiveness measurement system to continue to monitor program performance. The permanent system should conform to the structure, process, and quality requirements discussed in chapter 3.

Limitation of an ad hoc system

Although an ad hoc system should also be well designed, it is subject to limitations that would not otherwise affect a permanent effectiveness measurement system. For example, performance indicator data often has to be developed or reconstructed after the fact. For example:

> One service objective of a transit system is to provide on-time performance. If this data has not been previously collected, it would be inappropriate for the program results review team to expect program personnel to willingly duplicate past service performance. Instead, the review team may have to rely on the memories and

perception of the transit system users. Although more current data would be better for a permanent effectiveness measurement system, this surrogate approach may be the most appropriate under the circumstances.

A second limitation is the use of sampling to measure results. If original data is available, the review team will typically apply sampling techniques to gather the data. Under proper conditions, sampling can provide an acceptable level of data reliability, although it cannot duplicate the precision of an ongoing data collection system. For example:

> A vocational training program does not have a system for tracking former program participants to determine current employment status. Despite this limitation, an ad hoc system was designed to locate a random sample of former participants by using 3-year-old registration data. Although some participants could not be located, a sufficient number were found to permit a useful determination of program effectiveness.

A final limitation is the scope of the data required for an ad hoc system. For programs with multiple locations, the development of original data, even with sampling techniques, can consume considerable resources. As an alternative to developing data for all locations, the ad hoc system will generally rely on data taken from selected locations believed to be representative. To ensure the integrity of this approach, care should be exercised in selecting representative locations.

If limitations affect findings generated by an ad hoc measurement system, these limitations and the extent of their impact should be cited when discussing or reporting the effectiveness findings.

SCOPE OF AN AD HOC SYSTEM

The scope of an ad hoc system depends on the circumstance that necessitates its use. The most extensive scope is required when management has failed to establish a system or the system proves to be totally inadequate. Under these circumstances, the scope should include all objectives which are subject to review. In other situations, the scope of the ad hoc system is restricted. For example, when management's

system is partially acceptable, the ad hoc system need only measure those objectives that were previously omitted or improperly measured.

DEVELOPING AND IMPLEMENTING AN AD HOC SYSTEM

When an ad hoc system is needed, the review agency should solicit program management's cooperation and involvement. The involvement of program personnel in the design of an ad hoc system not only provides a valuable source of program knowledge but also enhances the development of a mutually acceptable system.

The development and use of an ad hoc system involves the following activities:

--Studying the feasibility of developing and implementing an ad hoc system.

--Designing the system.

--Collecting the data required by the system.

--Determining the level of program effectiveness.

Performing a feasibility study

If the achievement of certain program objectives is either not measured or inadequately measured, the review team should determine the feasibility of developing and implementing an ad hoc measurement system. This feasibility study should address those factors that may invalidate or significantly compromise the usefulness of data that can be generated by an ad hoc system.

A similar, yet less formal, assessment was made during the prereview planning stage. At that time, the review agency questioned the amenability of a program or selected program objectives to review based on the general program environment. The five conditions presented in Table 3 (see p. 27) should be reconsidered in greater detail during this feasibility study.

The review team should particularly consider the appropriateness and availability of performance indicators, data sources, and performance standards. In addition, the reasons provided by program management for not previously measuring effectiveness should be carefully assessed.

After weighing these factors, the review agency must decide the general feasibility of developing and implementing an ad hoc system. If such a system is determined to be inappropriate, the review agency will have to (1) terminate further work relative to the objectives for which the ad hoc system was being considered and (2) prepare a justification statement. If the feasibility study, on the other hand, indicates that an ad hoc system can be developed to generate reliable data, the review agency in conjunction with program personnel should undertake the actual design of the ad hoc system.

Designing the ad hoc system

The design of the ad hoc system must incorporate the same structure and process features discussed in chapter 3. In addition, the designers of the system must consider the same qualitative factors that are indicative of an acceptable system. In particular, the indicators should be valid and sufficient, the standards should be appropriate, the data collection techniques should be acceptable, and the data sources should be reliable and unbiased. As discussed previously, some deviation from these qualitative factors may be acceptable, depending on the affect such a deviation has on the effectiveness findings.

In the ideal situation, management will provide staff to assist in the development of the ad hoc system. In these cases, management's agreement as to the appropriateness of the system can generally be readily obtained. If assistance is not provided, the review team should solicit management's endorsement of the proposed ad hoc system. If the system is not acceptable to program management, the review team may be able to modify the system to gain management's endorsement without compromising the system's adequacy. If this approach fails, the review agency's actions will be determined by the conflict resolution agreements established during the pre-review planning stage; specifically, the review agency can (1) assert the ad hoc system without program management's endorsement, (2) rely on a higher moderating authority to resolve the conflict, or (3) terminate further work and prepare a justification statement.

The following example illustrates the process of designing an ad hoc system.

> During a program results review of a county program to treat alcoholic employees, the review team determined that an ad hoc system was necessary to measure the program's effectiveness because

management did not have an effectiveness measurement system. The review team discussed the situation with program management and determined that the development and implementation of an ad hoc system was feasible.

Although the program manager could not spare staff personnel to assist in the development of the ad hoc system, she asked to be kept informed.

To develop performance indicators and standards, the review team researched publications of other alcoholic treatment programs and Government health agencies. Additionally, officials of Federal and State treatment programs were interviewed. The research led to the identification and development of a set of indicators and standards that could be applied to the program under review. Following this work, the review team considered the problems of identifying data collection techniques and data sources. After addressing these problems, they discussed a proposed ad hoc system with the program manager.

Based on this discussion, a mutually acceptable system, as outlined in Table 4, was used to calculate program effectiveness.

Table 4

EXAMPLE OF A MUTUALLY ACCEPTABLE

AD HOC MEASUREMENT SYSTEM

I.	STRUCTURAL COMPONENTS:	
	A. Performance indicators:	1) Percentage of program participants reducing alcoholic consumption (divided into categories depending on the amount consumed).
		2) Percentage of program participants returning to the treatment program after being released.
	B. Data source:	1) Former program participants.
		2) Program files.
	C. Performance standard:	Results achieved by a comparable State-administered alcoholic treatment program.
II.	PROCESS ACTIVITIES:	
	A. Data collection process:	1) Confidential interviews with former program participants.
		2) Review and synthesis of information contained in program files.
	B. Comparison process:	Comparing the status of the county's performance indicator data with the comparable State statistics.

Collecting the data

After designing and deciding to implement the system, the necessary performance indicator data must be collected.

The data can be collected by either the review team or program staff. If program staff is used, the review agency should monitor the collection process to ensure data integrity. The general requirements discussed on pages 37 through 39 should be followed. In addition, the review agency may find it appropriate to independently verify the accuracy of this data by tracing selected data back to the original sources.

Determining effectiveness

After collecting the performance indicator data the review team should determine the level of program effectiveness. This involves comparing the actual level or status of the performance indicators with the performance standards. If, in aggregate, the comparison shows that actual program results fall below expectations, the review team should analyze the causes of ineffectiveness. (See ch. 7.) If the program is found to be achieving a satisfactory level of program results, the review team should obtain whatever supplemental information is necessary (see ch. 8) and communicate the review findings to the appropriate government officials (see ch. 9).

CHAPTER 7

IDENTIFYING CAUSES THAT INHIBIT PROGRAM EFFECTIVENESS

If a program is not achieving a desired level of results, the review team must identify the causes that inhibit program effectiveness. Based on these causes, appropriate recommendations for improving future program performance can be formulated.

This chapter discusses

--the factors that affect program performance,

--a search process for identifying potential causes of ineffectiveness, and

--a hierarchy of methods to validate the reliability of the identified causes.

FACTORS AFFECTING PROGRAM PERFORMANCE

Factors that affect program performance can be categorized into three general areas. The first category involves the adequacy of a program's operational plans, policies, procedures, and controls. The second category relates to the limitations imposed by program inputs. The last category consists of the compatibility of the program to its external environment. The following subsections discuss these three general areas together with explanatory examples.

Operational factors

Program plans, policies, procedures, and controls are designed specifically to direct program operations toward the attainment of desired objectives. If program objectives are not satisfactorily achieved, the cause may be rooted in an inadequate operating structure. For example, operating plans may be vague or nonexistent, policies and procedures may improperly reflect program objectives, or controls may be too weak to assure that otherwise sound procedures are being correctly followed. Any of these inadequacies can severely diminish a program's effectiveness.

The following examples demonstrate the effect of inadequate planning, direction, or control on the attainment of program results:

Inadequate planning

Program - Federal Highway Safety (Construction) Program

Objective - To reduce death, injuries, and property damage caused by traffic accidents on the Nation's highways

The review staff found that the highway death rate had declined over the past 9 years, but not to an acceptable level. While reviewing the program's planning process, the review staff noted that the information available for making project selection decisions was inadequate. Highway accident data had not been analyzed to determine the most hazardous locations and, as a result, inventories of cost-effective projects had not been developed to establish priorities. This lack of a systematic and informed planning function allowed projects to be funded in low-hazard locations, thus preventing the optimum safety benefits that could result from the construction program.

Inadequate policy or procedural direction

Program - Small Business Procurement and Contract Program

Objective - To ensure a fair proportion of total government purchases and contracts are placed with small business concerns

Based on a review of records, interviews with program personnel, and questionnaires prepared by program participants, the review staff found that some vendors and contractors receiving program benefits did not meet the eligibility requirements intended by the authorizing legislation. A further search revealed that the program had neither established a policy defining the qualification requirements for small business concerns nor designed procedures to check the size or need of program participants. As a result, program benefits were being misdirected, thereby diminishing the level of program effectiveness.

Inadequate control

Program - Boating Safety Program

Subprogram - Penalty Assessment Program

Objective - To stimulate safe boating practices through the deterrent effects of penalty assessments

Although specific guidelines establishing monetary penalties for the violation of safe boating laws had been established, subprogram management opted to follow a more lenient policy. They allowed individuals violating safe boating laws to attend educational workshops in lieu of a penalty assessment. Higher level program authority never monitored or reviewed this deviation from established procedure. During the program results review, a significant number of second- and third-time offenders were noted, indicating that the lenient actions taken by subprogram personnel were not effective. Since acceptable guidelines and procedures existed, the lack of adequate program control diminished the potential effectiveness of the penalty assessment program.

Program inputs

Every publicly funded program must operate within an environment confined by program inputs. These inputs include not only budget and staff limitations, but also various legal or regulatory requirements that affect program operations. Occasionally, these resource limitations or compliance requirements can constrain program effectiveness.

Review staff should be cautioned that an increased budget or additional staff does not automatically imply increased benefits or results. Before such a conclusion can be drawn, the review agency must demonstrate that additional benefits will in fact result from the proposed increase in resources.

The following examples illustrate the constraining effect program inputs can have on the attainment of program objectives.

Resource limitations

Program - Highway Beautification Program

Subprogram - Junkyard Screening Program

Objective - To remove or screen roadside junkyards

A comparison between the number of junkyards removed or screened from view and the total inventory of roadside junkyards disclosed that little success had been achieved since the program was initiated. Even the few screens that were constructed were determined to be inadequate and poorly maintained. Through interviews and observations, the review staff found the low program effectiveness was due to a general lack of emphasis and support. For example, minimal

staff time was allotted for this subprogram relative to other Highway Beautification subprograms. Furthermore, insufficient funds were available to accomplish the screening objective. By developing cost data for building and maintaining junkyard screens, the review staff was able to demonstrate the level of program effectiveness that could be achieved with commensurate funding increases.

Compliance requirement

Program - Senior Citizens' Nutrition (Grant) Program

Objective - To provide meals to senior citizens who cannot otherwise afford nutritious meals

Interviews with program participants revealed that social rather than economic needs prompted many senior citizens to eat in the program's centralized dining facilities. As a result individuals not falling into the legislatively intended target group were being provided meals, thus denying program benefits to otherwise eligible senior citizens. This condition was traced to a compliance requirement imposed on the grantee. The requirement, which was intended to reduce the welfare stigma, prohibited program personnel from soliciting financial data from program participants. Although this compliance requirement was being satisfied, the program was constrained from satisfactorily accomplishing its program objective.

External environment

Factors external to a program can also inhibit the accomplishment of program objectives. Such factors are generally outside the program manager's control and do not hinder the production of the program outputs. Instead, these external constraints intervene to prevent the program outputs from producing the desired program result or societal change.

A fine line exists between attributing program ineffectiveness to an external cause or to inadequate program planning. The distinction lies primarily in the degree of knowledge or predictability of the external factors coupled with managerial freedom to redirect program resources toward desired goals. If an external factor was known or predictable and the program manager had the flexibility to avoid its influence by changing the program's operating strategy, the cause of reduced effectiveness may be more aptly identified with inadequate program planning. Regardless of this distinction, the review team should primarily be concerned with whether or not program effectiveness can be improved.

The following examples involve (1) a program that can change its operating strategy to avoid the influence of the external constraint and (2) a program that is so severely constrained that termination is the most practical solution.

Alternative operating strategy

Program - High School Vocational Education Program

Objective - To train youth for vocational employment

The review staff found that the high schools in a major metropolitan area offered a diverse vocational education program. A random sampling of program participants, however, revealed that training in two fields, carpentry and printing, had not resulted in posttraining employment. Interviews with program graduates and other individuals, knowledgeable of the job opportunities in the respective fields, disclosed the following (external) factors that hinder postgraduation employment opportunities.

--Carpentry: a change in the union-sponsored apprenticeship examination emphasized mathematical skills not covered in the vocational education course.

--Printing: the introduction of semiautomated printing equipment requiring special operational skills rendered vocational graduates unemployable in all but the smallest printing shops.

Although these external factors inhibited postprogram effectiveness, a minor adaptation in the program's operating environment could improve future program performance. For example, the carpentry course could easily include the required mathematic skills in its classroom training; and the printing course could either acquire the necessary equipment or possibly arrange on-the-job training opportunities with local printing establishments that have the new printing equipment.

Program termination

Program - Imported Fire Ant Program (Federal)

Objective - To eradicate, suppress, or control imported fire ants

The performance data disclosed little change in the total land area infested by imported fire ants. The review agency found two external causes that limited the program's effectiveness: (1) the ability to effectively eradicate or suppress imported fire ants was due to (external) restrictions on known chemical extermination agents and (2) Federal control activities were found to primarily duplicate similar (external) State efforts. Since the external environment either prohibits the eradication of fire ants or already provides considerable fire ant control protection, the review staff recommended the program either be terminated or reduced to fund only those control activities not performed by State agencies.

SEARCH FOR CAUSES OF INEFFECTIVENESS

Keeping in mind the previous factors that affect program performance, the review team can begin the search for specific causes that prevent a particular program from satisfactorily achieving its intended results. This search process involves soliciting leads from individuals knowledgeable of the program and its environment as well as independently examining or evaluating available information sources. The review team must both ask and answer "why" questions; that is, questions that will offer explanations as to why a program is operating below its expected performance level. The following subsections discuss specific search activities through which possible causes of ineffectiveness can be identified.

Interview program personnel

Both program managers and staff personnel may be aware of actual or potential problems that inhibit program performance. Review staff should not only ask specific "why" questions but also be alert for unsolicited comments or leads.

Program managers may be already aware of major program constraints. When informed of an unfavorable level of program performance, program managers will usually offer some type of explanation. The review staff must determine the basis and reasonableness for management's explanations and ascertain why corrective action has not previously been taken. In many instances, program managers will defend their performance by only identifying constraints that are beyond their ability to control (for example, program inputs and external factors).

Program personnel below the management level may also be able to provide possible leads. For example, they may be able to identify specific problems that hinder their own

performance such as confusing procedures, insufficient training opportunities, or inadequate equipment.

Review previously collected performance data

The performance data that was reviewed and evaluated during the determination of effectiveness process may also provide clues to ineffective performance. For instance, data that was compiled from case files may indicate a consistent error possibly caused by an inadequate procedural or control function.

Review previous audit reports or evaluation studies

Previous audit reports or evaluation studies that involve a determination of program effectiveness provide an excellent source of information. Program personnel may provide these documents, or inquiries can be made of the audit or evaluation groups that have jurisdiction over the program under review. If program audits or evaluations had been performed, the review staff should ascertain the status or current appropriateness of the previous findings and recommendations.

Review program plans, policies, procedures, and controls

Review in depth the program's plans, policies, procedures, and controls. An objective, unindoctrinated examination may isolate problems or inadequacies that neither program management or personnel had previously considered.

Observe program operations

Again, an objective observation of program activities may reveal improprieties or inconsistencies that had previously gone unnoticed. For example, to expedite the number of food stamp applications, a program claims reviewer may overlook a time-consuming eligibility verification procedure that if followed would reduce the misappropriation of program funds to ineligible recipients.

Interview program clients or beneficiaries

The individuals for whom a program's benefits are directed may also be able to identify program inadequacies. This is especially true if the program has a service, usage, quality, or equity condition for which the perceptions of

program clients are critical for program success. For example, eligible citizens may be able to identify reasons why a public recreation facility is underutilized, or program clients may be able to isolate service problems relating to a public transit program.

Assess the sufficiency of program resources

Program objectives may be too ambitious for the resources allotted to the program. If this is suspected, the review staff should assess whether or not program resources are sufficient to accomplish the intended result. For example, a motor vehicle safety inspection program staffed with only one inspector for every 100,000 vehicles is probably inadequate to ensure compliance through either enforcement or deterrent efforts.

Analyze program trends

The maturity of a program coupled with its performance record may explain why a program is currently not achieving a satisfactory level of effectiveness. Few programs are instantaneously effective. An analysis of a program's performance trend may show that the level of effectiveness is increasing over time.

Likewise, the analysis of a program's performance trend may indicate a deviation that can be related back to a change in the program's operating environment. For example, abnormal weather conditions may affect the use of an outdoor recreational program.

Examine the program's external environment

The review team should also consider external factors that may keep a program from affecting the desired result or societal change. The development of such factors may require both imagination and insight. Examples of external factors that can influence the effectiveness of a vocational education program are unemployment rates by vocational categories, union restrictions, and teacher tenure rights.

Interview other individuals knowledgeable of the program or its environment

Individuals who are neither program employees or beneficiaries may also be familiar with the program's operations or external environment. If individuals are experts in a particular subject area, such as health, they may be knowledge-

able of the problems that may affect the achievement of program objectives. Other knowledgeable individuals include former program employees or former beneficiaries who may be able to more candidly discuss program problems than current employees or beneficiaries.

VALIDATING THE POSSIBLE CAUSES

The previous section describes a search process for identifying possible causes of ineffectiveness. Before a possible cause can be reported as actual, the review staff will have to substantiate its validity. The extent of validation work must be sufficient to convince the appropriate government officials to initiate corrective action. Key variables that affect the selection of a validation process are how obvious the cause is and its acceptance by program management.

The following four subsections discuss a hierarchy of validation schemes that are based on increasingly sophisticated cause-and-effect evidence.

Agreeable to program management

Some causes are blatantly obvious to both the review staff and program management. In such cases, documenting the extent to which the causal factors inhibit program performance is generally unnecessary. The review staff should first inform program management of the identified cause and solicit their agreement. If management disagrees, the review staff should consider the basis for their disagreement and, if appropriate, develop whatever additional evidence is necessary to convince the review requestor or appropriate oversight authority of the validity of the identified cause.

In the following example, the cause of ineffectiveness was agreeable to program management and thus required minimal formal verification.

Program - Department of Defense Inactive Item Program

Objective - To eliminate unneeded, inactive items from the supply system

The review disclosed that both the number and supply cost of inactive items had grown despite program efforts. Because the users of defense-supply items, as well as the storage locations, are so numerous and diffused, the review staff examined

the process for collecting and disseminating supply and item usage information. Based on their observations, the lack of an integrated information system was found to be the primary cause for ineffective performance. Without further validating the causal relationship between the information system and program performance, the review staff recommended and program management agreed that an adequate and cost-effective computer system should be established to provide prompt and complete supply and user information.

Increasing the scope

Although a problem may have been identified during the search process, insufficient causal evidence may initially be available to ensure common agreement. In such cases, the volume or scope of supporting data may need to be enlarged to demonstrate the appropriateness of the previously identified cause. Such an enlargement may, for example, involve:

--Augmenting the number of review locations. (For example, an inappropriate practice found to exist at 2 of 10 program field offices may need to be demonstrated at 3 additional field locations.)

--Increasing the statistical reliability of previously generated causal data. (For example, the verification of a cause identified during preliminary interviews with a small number of program clients may require a larger or statistically significant sample size.)

Corroboration

Merely increasing the number of observations or opinions from a single source may be insufficient to convincingly demonstrate the validity of a previously identified cause. A more involved approach requires a corroboration of two or more sources. Examples of this type of corroboration are:

--An agreement between separate random samplings of program clients and program employees that the lack of scheduled bus service is the cause for low usage.

--The independent review of program procedures and the observation of personnel practices may similarly demonstrate vague or confusing procedural instructions.

--An analysis of performance trends coupled with an examination of a program's external environment

may show that the decrease in a program's job placement statistics directly coincides with local unemployment rates for the program target group.

Regression analysis

Regression analysis adds a more mathematically rigorous dimension to the validation process. Basically, it involves relating the effect of one or more identified causes, or independent variables, to the program result, or dependent variable. Such analysis requires that data relating to causes and program results be measurable, accurate, and available in a sufficient number of combinations (for example at different points in time or at a variety of locations).

The effective use of regression analysis requires some mathematical expertise. While the individuals conducting a review may not have the requisite expertise, they should be able to recognize the potential application of regression analysis and have access to specialists who can develop a proper mathematical model. An example follows.

Regression analysis

Program - Affirmative Fair Housing Marketing Program

Objective - To improve the racial balance of individuals with similar income levels in the same housing market area

By comparing the racial composition in housing projects subject to Affirmative Fair Marketing Program regulation with a control group of housing projects not subject to such regulation, the review staff concluded that the program had no measurable affect upon improving racial housing balance. To determine what does influence the racial distribution within single family subdivisions and multifamily projects, the review staff designed a multiple regression model to compare the interrelationship of the percentage of minorities in a project or subdivision with several independent variables thought to have a direct or indirect effect. The independent variables included

--income of the neighborhood,

--percent of minorities in the neighborhood, and

--presence or absence of a housing subsidy.

Based on the regression analysis, the location or neighborhood of the project was the strongest determinant of racial composition. In other words, the racial composition of the housing project reflects the prevalent minority composition of the neighborhood despite the Affirmative Fair Marketing Program's regulatory requirements.

SUMMARY

To properly identify the cause or causes that inhibit a program from satisfactorily achieving a program objective, the review team must know

--what factors affect program performance,

--how to identify possible causes of ineffectiveness, and

--how to substantiate the validity of a previously identified cause.

CHAPTER 8

OBTAINING SUPPLEMENTAL INFORMATION

The previous three chapters discussed the specific activities that are necessary to satisfy the basic objectives of a comprehensive program results review. Supplemental information, however, may also be required, particularly if it provides additional perspective for government decisionmakers. Although the discussion of supplemental information has been reserved for a separate chapter, it can generally be obtained or developed concurrent with the previous review activities.

The following three categories of supplemental information should be considered in conjunction with all program results reviews:

--The development of cost data.

--Management's prior consideration of lower cost alternatives.

--The impact of unintended program results.

DEVELOPMENT OF COST DATA

The relevancy of cost data is subject to some dispute. Many theorists believe that effectiveness should not be considered in relation to cost; such an endeavor is, instead, a function of efficiency analysis. The reasonableness of program costs, nonetheless, is an implicit program objective. Program managers and oversight bodies recognize the importance of cost-effective program performance. The following example illustrates this point:

> A National Park Service land acquisition program has an objective "to acquire privately owned property within the boundaries of certain national parks." Since this objective does not explicitly limit the price that can be paid for these lands, the level of program effectiveness could be expected to increase as the prices offered the property owners begin to exceed appraised values. Overriding this program strategy, however, is the implied objective of reasonableness. Program managers are aware that their program will be judged not solely on effectiveness

but on cost-effectiveness. Thus, private lands are only acquired when their sales prices reasonably reflect market values.

Two types of cost data are important to the requestors and users of program results review information. The first involves the actual cost of the program under review. The second relates to the cost that would be incurred to improve program effectiveness.

Actual program cost data

Program cost data can generally be obtained from operating budgets or other financial documents. If program budgeting is not used, the actual cost of a program may be lost in the myriad of activities undertaken by the agency in which the program is contained. If cost data cannot be obtained, the review team should document the reason.

Occasionally, the review team may find it appropriate to elaborate upon this basic program cost data. For example, if a program consists of numerous activities or involves a variety of locations, cost data that is properly categorized may provide additional insight as to the relative cost-effectiveness of the different activities or locations. Likewise, significant increases or decreases in cost data from one year to the next may be of interest, particularly if they parallel changes in annual program performance.

Program improvement costs

After the causes that inhibit effectiveness are identified, the review team can formulate appropriate recommendations. These recommendations must not only suggest alternative operating strategies that will eliminate or reduce the influence of the identified causes, they must also be reasonable; that is, the cost of implementing the recommendation must not exceed the benefits to be derived. If a recommendation causes a substantial increase in program costs, the review team should estimate and report both the costs and expected benefits. For example, a recommendation to increase the number of personnel conducting safety inspections should include estimates of both the additional salaries and the reduction in accidents that can be expected.

MANAGEMENT'S PRIOR CONSIDERATION OF LOWER COST ALTERNATIVES

The gold book definition of a program results review includes more than the conventional examination of effectiveness as presented thus far. The definition also includes a supplemental condition that requires the program results review team to determine whether program management "* * * has considered alternatives which might yield desired results at a lower cost."

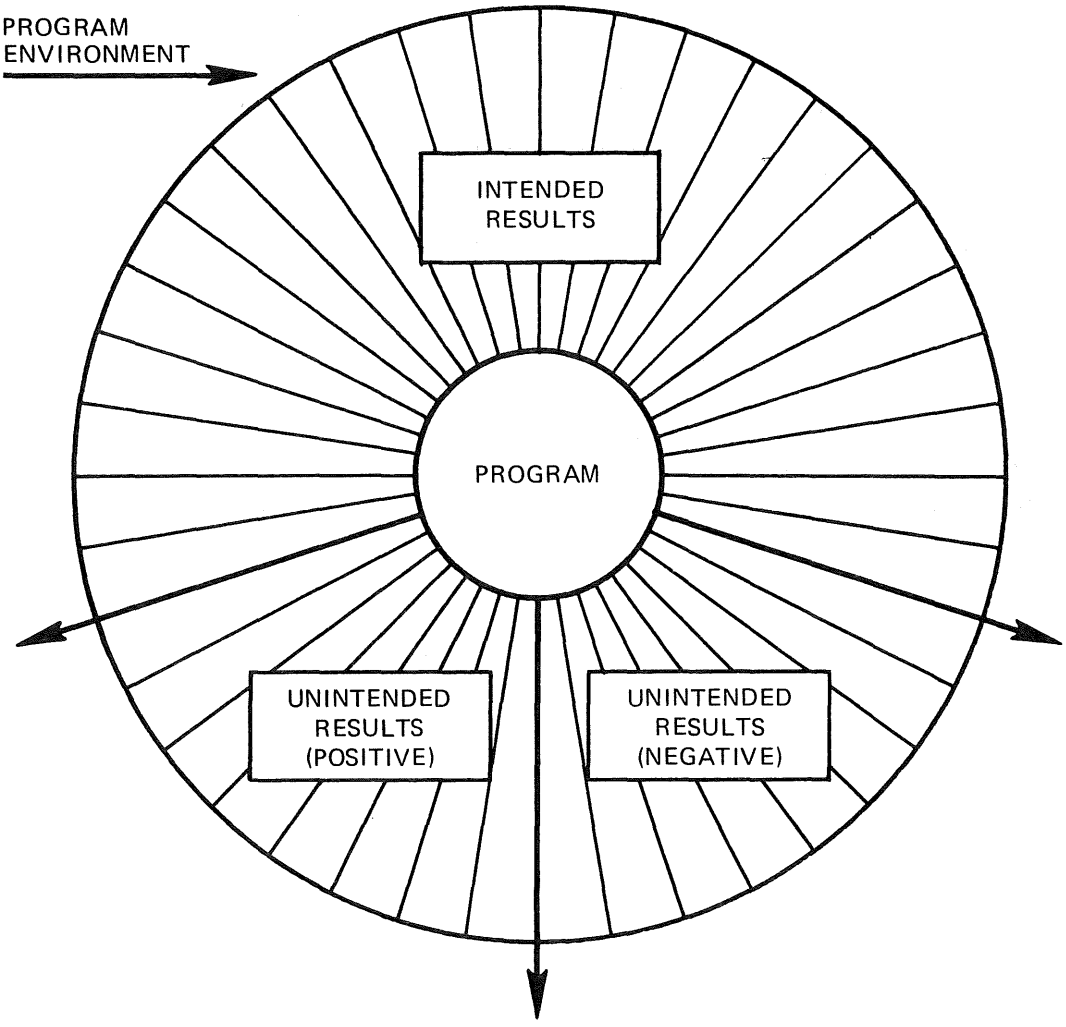
Unlike the economy and efficiency review objective, that requires an independent examination of all program operating activities, this condition only requires the review team to solicit and verify information provided by program management. If program management has not considered less costly alternatives, the review team should determine the reason such alternatives were not considered. If, on the other hand, less costly alternatives have been considered, the review team should determine how these alternatives were developed and whether program management plans to implement them.

IMPACT OF UNINTENDED RESULTS

A program can affect various changes on the society or environment in which the program operates. The review process thus far has only considered those changes or results that reflect the intent of a program objective. Other unintended results may also occur as program side effects. These unintended results can be either positive or negative depending on the benefits or disbenefits they engender. Together, all the results attributable to a program constitute that program's total impact on society (see Figure 3).

FIGURE 3

TOTAL PROGRAM IMPACT



Although the program results review is primarily concerned with desired or intended results, the review team should also be alert for unintended results. If they are observed or identified, their relative significance should be determined. This decision will often be subjective. To formally determine their significance would entail the development of an effectiveness measurement system comparable to that used to measure intended results. If based on the judgment of the review team the unintended results are sufficiently significant, their existence and estimates of their impact on society should be reported together with the other review findings.

The following examples illustrate the nature and impact of unintended results:

A positive unintended result

A county program has the objective to provide meals to indigent elderly citizens. A program results review found, however, that restrictions against the development of financial data caused the program to serve meals to not only financially deprived elderly citizens as intended by the program objective, but also to socially deprived elderly citizens. Although the final report included a determination of program effectiveness relative to the mandated program objective, it also discussed the extent of social benefits derived by this unintended program result.

A negative unintended result

A training program with an objective to provide individuals with employment skills was found to be highly successful in placing graduates in good jobs. The review team observed, however, that program graduates were displacing and rendering unemployable older workers in the same skills areas. The extent to which this negative unintended result is attributable to the program may cause government decisionmakers to reconsider the value of the training program in relation to its total societal impact.

Total program impact

A program was established by a city police department to decrease the number of accidents,

deaths, and injuries caused by drunk drivers. The program involved reassigning police officers to locations known to have a high incidence of drunk drivers. This program caused the following changes:

I. Intended results
--The number of accidents, deaths, and injuries caused by drunk drivers decreased as intended.

II. Unintended results - positive
--The number of accidents, deaths, and injuries caused by nondrunk drivers decreased.

III. Unintended results - negative
--Crime rates in other parts of the city increased due to the reassignment of the police.

--An additional burden was placed on the municipal court from the increase in drunk driving violations.

--Harassment complaints increased due to the number of nondrunk drivers who were stopped and subjected to sobriety tests.

CHAPTER 9

COMMUNICATING THE REVIEW FINDINGS

The last major review activity involves communicating the review findings to the appropriate government officials. Written reports are the normal method for communicating review findings. Oral briefings, however, may also be required or appropriate. Regardless of the type of communication, the quality and content of this activity is governed by the gold book reporting standards and the specific review objectives pursued.

The gold book reporting standards establish the quality requirements that apply to all audit reports. The standards specify such desirable characteristics as timeliness, conciseness, balance, and objectivity. In addition, the standards emphasize the appropriateness of such activities as obtaining the comments of responsible program officials and distributing the final report to all interested parties.

The actual content of a program results review report depends on the scope of the review assignment. The remainder of this chapter discusses the types of findings, conclusions, and recommendation that might result from a comprehensive program results review. Briefly, these involve:

--A determination of whether the desired results are being achieved and how this determination was made.

--The adequacy of management's system for measuring effectiveness and general recommendations for improving the system, if appropriate.

--The identification of causes that inhibit the intended level of program effectiveness together with appropriate corrective recommendations.

DETERMINATION OF EFFECTIVENESS

The level of program effectiveness is based on the findings generated by either management's effectiveness measurement system (see ch. 5) or an ad hoc system (see ch. 6). Regardless of the system used, the review team should report whether the program is achieving its desired results.

The review agency is not required to express an opinion as to the absolute effectiveness of a program. Such a

definitive opinion would only be possible if program objectives were directly quantifiable and the program was charged with achieving a specific level of program results. In the absence of this precision, the review agency must qualify the effectiveness determination based on the system used. Therefore, a complete description of the effectiveness measurement system, including the performance indicators, performance standard, data sources, and data collection techniques, should be included in the final report. If appropriate, the reason for using this system should also be included.

If effectiveness cannot be determined, the review agency should discuss the reasons that prevent its measurement. Some of the problems that diminish the ability or desirability of measuring effectiveness were presented in Table 3.

ADEQUACY OF MANAGEMENT'S EFFECTIVENESS MEASUREMENT SYSTEM

Based on the requirements described in chapter 5, the review team can assess both the adequacy of management's effectiveness measurement system and whether it is sufficiently comprehensive to measure all program objectives. When deficiencies are noted, the review agency should recommend appropriate remedial action.

Table 5 shows the possible conclusions and recommendations that might result from this assessment.

Table 5

CONCLUSIONS AND RECOMMENDATIONS

RESULTING FROM THE ASSESSMENT

OF EFFECTIVENESS MEASUREMENT SYSTEMS

CONCLUSIONS:	RECOMMENDATIONS:
1) System is adequate and comprehensive.	1) None.
2) System is adequate, but does not measure all program objectives.	2) Program management should expand system to measure effectiveness relative to all program objectives.
3) System is inadequate, but correctable.	3) Program management should modify system to correct the deficient performance indicator, data source, performance standard, or collection techniques.
4) System is totally inadequate.	4) Program management should design a new measurement system that incorporates the appropriate quality characteristics.
5) System is nonexistent.	5) Program management should design an acceptable system to measure program effectiveness.

CAUSES OF INEFFECTIVENESS

If a program is not achieving a desired level of results, the report should identify the causes that prevent effective performance. The activities described in chapter 7 provide a process for both identifying and validating these causes.

The review team should formulate appropriate recommendations to either eliminate or reduce the influence of these causes. Ideally, these recommendations will constitute minor

changes in the program's operating strategy. The review team must, however, be sensitive to both the cost of the recommendation and the nature of the cause. If the cost of the recommendation exceeds the benefits or if the cause is so significant that its influence cannot be diminished, the most appropriate recommendation may be a reduction or termination of program operations.

If various alternative operating strategies can eliminate or reduce the causes of ineffectiveness, the final report should present each alternative together with its anticipated benefits and costs. From this range of alternatives, the appropriate government decisionmakers can select the one alternative that best reflects their program policy expectations.

APPENDIX I

COMPREHENSIVE FLOW CHART

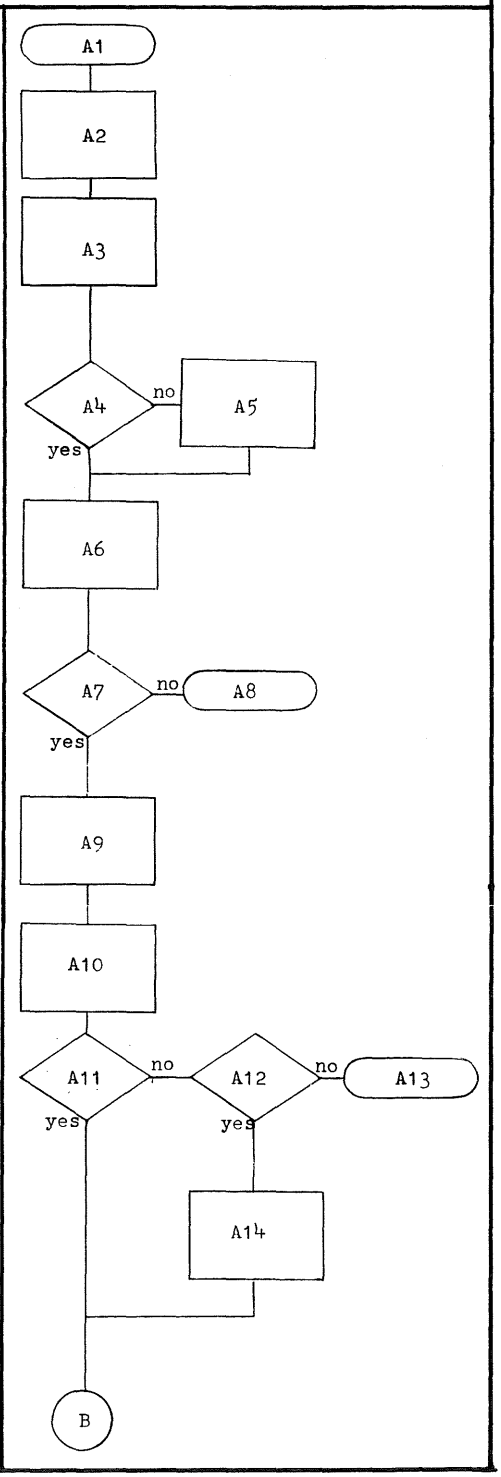
The following flow chart depicts the interrelationship of the major program results review activities described in chapters 4 through 9. Where appropriate specific page references are also provided.

Flow chart	Review function	Chapter
A	Prereview familiarization and planning activities	4
B	Assessing management's effectiveness measurement system	5
C	Using an ad hoc system to measure effectiveness	6
D	Identifying causes that inhibit program effectiveness	7
E	Obtaining supplemental information	8
F	Communicating the review findings	9

APPENDIX I

A: PRE-REVIEW FAMILIARIZATION AND PLANNING ACTIVITIES

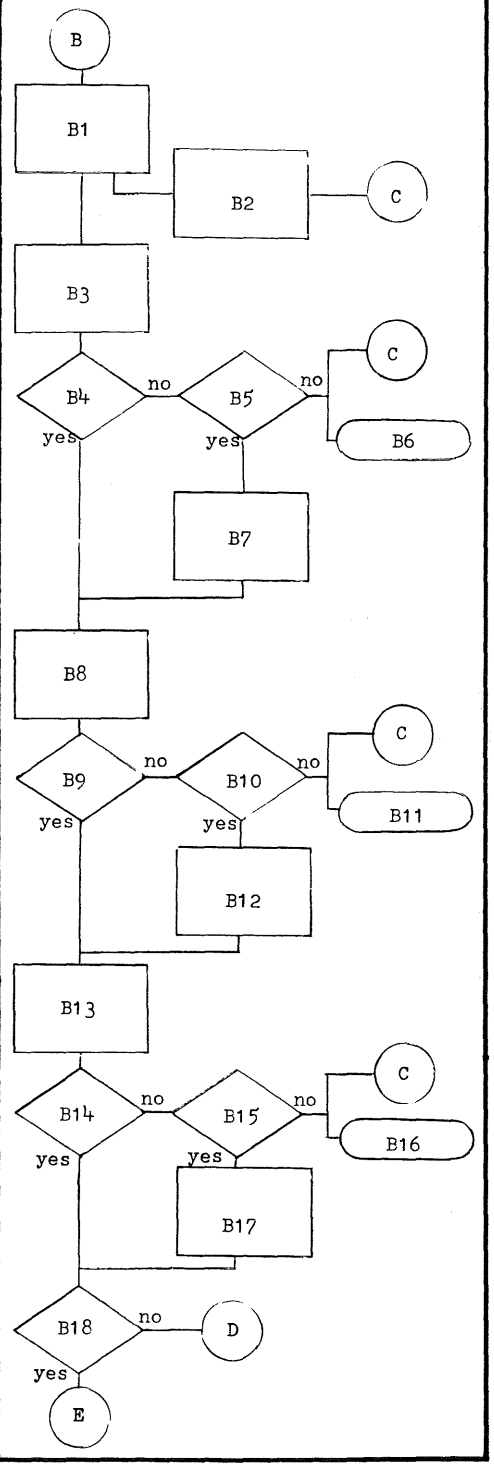
	Page:
A1. Selection of program for review.	
A2. Receipt of assignment request or RFP.	
A3. Assignment clarification activities:	
-- assignment objective.	19
-- basic review and reporting requirements.	20
-- program objectives.	21
-- potential problems.	21
A4. Are the assignment objectives and requirements clearly understood?	
A5. Obtain additional clarification from the review requestor.	
A6. Conduct a preliminary survey to	
-- verify program objectives.	24
-- become familiar with program operations.	24
-- identify the existing system for measuring program effectiveness.	25
A7. Based on preliminary program knowledge and review agency capabilities, should program results review be pursued?	26
A8. Terminate further involvement; if necessary, prepare written justification.	
A9. Select an appropriate review strategy.	28
A10. Prepare work plan and/or bid proposal outlining proposed strategy, timetables, and resource requirements.	29
A11. Is work plan (bid proposal) acceptable?	
A12. Can work plan (bid proposal) be modified to be made acceptable?	
A13. Terminate further involvement.	
A14. Modify work plan (bid proposal) as necessary.	



APPENDIX I

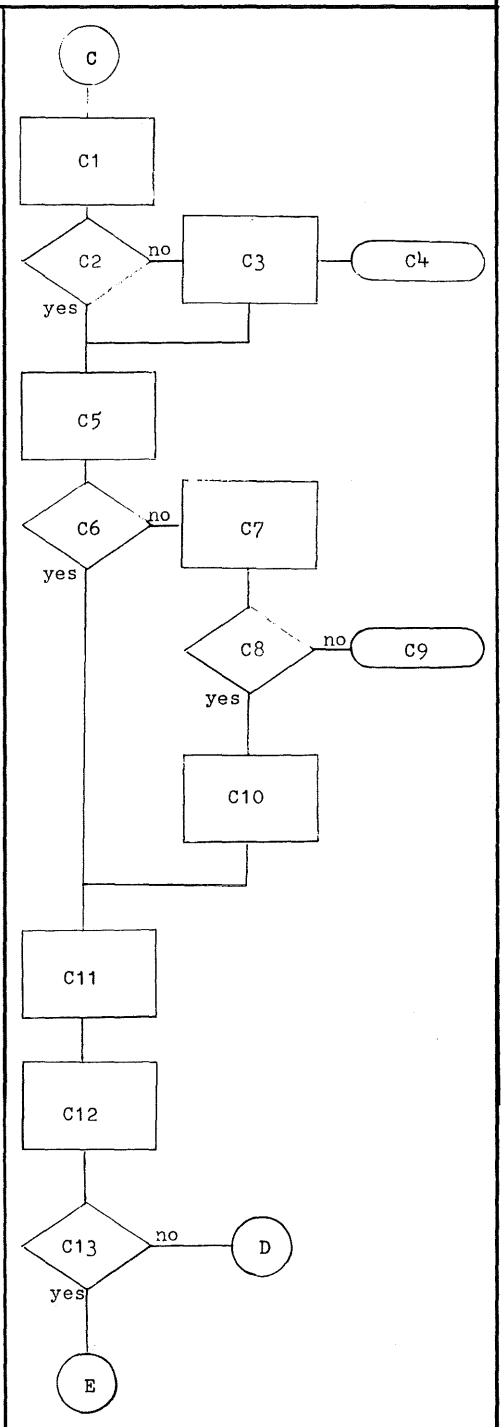
B: ASSESSING MANAGEMENT'S EFFECTIVENESS MEASUREMENT SYSTEM

	Page:
B1. Identify and document management's system for measuring program effectiveness.	31
B2. Identify program objectives not subject to effectiveness measurement; consider feasibility of an ad hoc system for these objectives.	31
B3. Assess the validity and sufficiency of the performance indicators.	32
B4. Are the performance indicators valid and sufficient?	
B5. Does management agree with the identified deficiency?	34
B6. Without the approval or authority to independently develop an ad hoc measurement system, terminate further work relative to the deficient condition and prepare an appropriate justification statement.	35
B7. Modify or monitor the modification of management's effectiveness measurement system.	35
B8. Assess the accuracy of the performance indicator data.	35
B9. Do the data sources and collection techniques provide reliable and unbiased data?	
B10. Does management agree with the identified deficiency?	40
B11. Same as "B6" above.	
B12. Same as "B7" above.	
B13. Assess the appropriateness of the performance standards.	41
B14. Are the performance standards appropriate?	
B15. Does management agree with the identified deficiency?	44
B16. Same as "B6" above.	
B17. Same as "B7" above.	
B18. Based on either the existing or modified effectiveness measurement system, is the program achieving the desired level of program results?	44



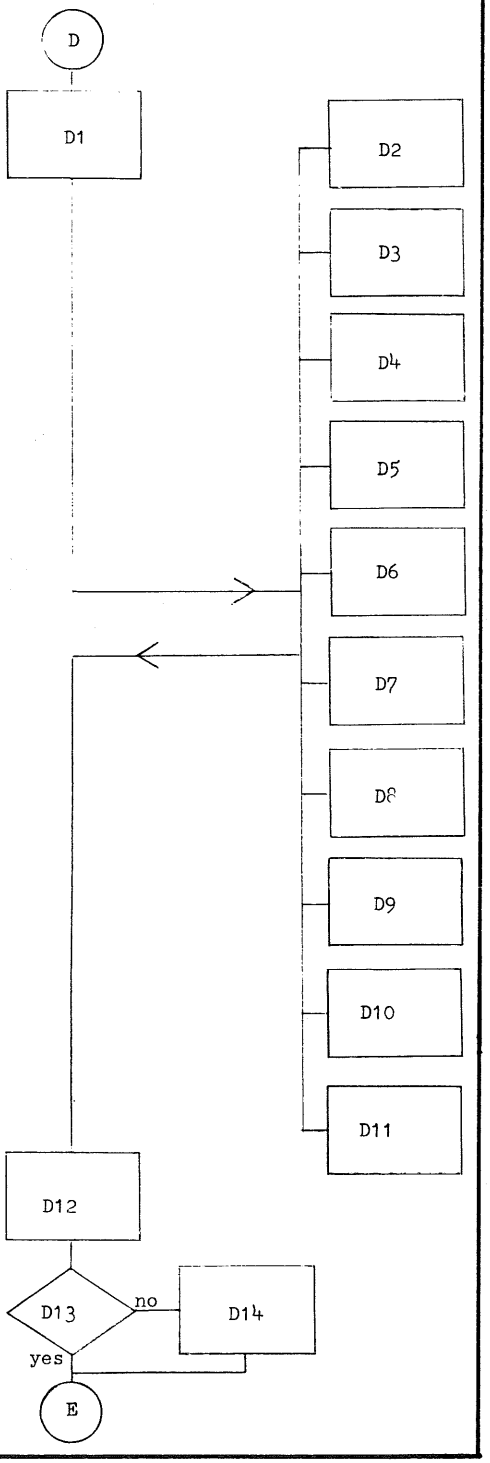
C: USING AN AD HOC SYSTEM TO MEASURE EFFECTIVENESS

	Page:
C1. Conduct a feasibility study.	48
C2. Can an acceptable ad hoc measurement system be developed?	
C3. Document reason(s) an ad hoc system cannot be used to determine program effectiveness.	
C4. Terminate further efforts to measure that program objective.	49
C5. Design the ad hoc system.	49
C6. Is the ad hoc system acceptable to program management?	49
C7. Identify the cause of disagreement.	
C8. Based on the previously established conflict resolution agreements, can the review agency proceed to use the ad hoc system?	
C9. Terminate further work relative to the use of an ad hoc system.	
C10. Revise the ad hoc measurement system, if required.	
C11. Collect the necessary performance indicator data.	52
C12. Determine the level of program effectiveness.	52
C13. Is the program achieving its intended results?	



APPENDIX I

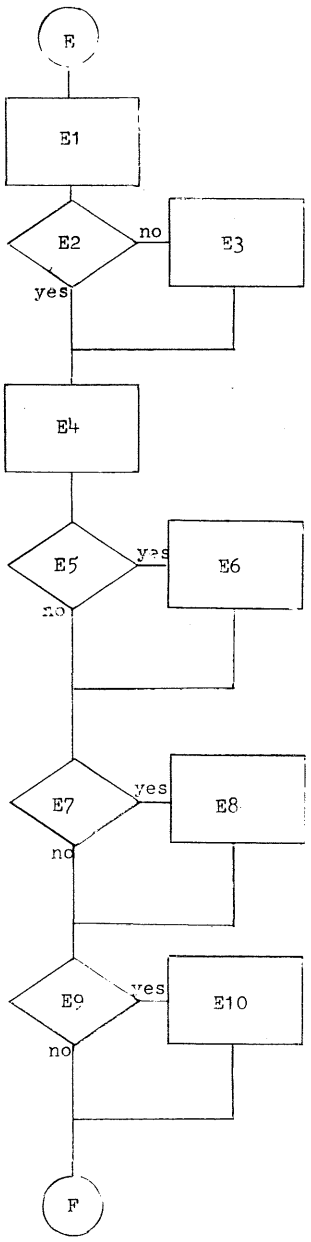
D: IDENTIFYING CAUSES THAT INHIBIT PROGRAM EFFECTIVENESS



APPENDIX I

E: OBTAINING SUPPLEMENTAL INFORMATION

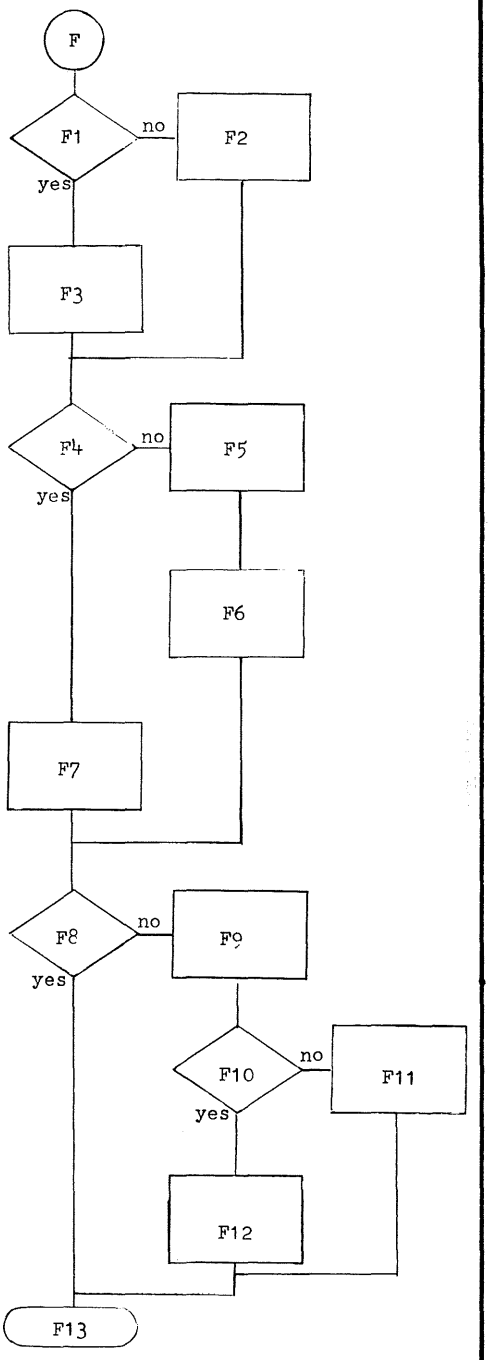
	Page:
E1. Solicit program cost data.	66
E2. Can program cost data be obtained or developed?	66
E3. Document the reason this data cannot be obtained or developed?	
E4. Develop or estimate the cost of eliminating or reducing the influence of the validated causes of ineffectiveness.	66
E5. Would significant costs be incurred to improve program effectiveness?	
E6. Estimate the benefits that would be derived from these costly proposals.	
E7. Has management considered alternative operating strategies that might yield the desired results at a lower cost?	67
E8. Document the status of these alternatives.	67
E9. During the review, were potential unintended results (i.e. side-effects) observed or identified that might be attributable to the program?	69
E10. Determine the significance of these unintended results.	69



APPENDIX I

F: COMMUNICATING THE REVIEW FINDINGS

	Page:
F1. Was the review able to establish a definitive determination of program effectiveness?	71
F2. Document and report the reason(s) that hindered the determination of program effectiveness.	72
F3. Report the level of effectiveness and the means (i.e. effectiveness measurement system) used to determine it.	72
F4. If management's effectiveness measurement system was used to determine effectiveness, did it operate adequately?	72
F5. Document and report the nature of the inadequacy.	
F6. Propose recommendations to correct the deficient condition.	73
F7. Report the verified adequacy of management's effectiveness measurement system.	73
F8. Was the program found to be achieving its desired results?	
F9. Document and report the causes that inhibit increased effectiveness.	73
F10. Can cost-effective recommendations be proposed?	
F11. Report the excessive costs or other disbenefits that would be necessary to improve effectiveness; recommend either program reduction or termination, if appropriate.	74
F12. Formulate and propose recommendations or alternatives to improve future program effectiveness.	74
F13. Issue report to appropriate government decision-makers.	



(91184)