

085075

11



095085

STATUS & ISSUES



Federal Program Evaluation

OCTOBER 1978

PAD-78-83

9105-1116

For sale by the Superintendent of Documents, U.S. Government
Printing Office, Washington, D.C. 20402

FOREWORD

In the 1970s, there has been an increased interest in program oversight and accountability. Public officials, legislators, and the public are increasingly asking whether

- programs are being implemented as intended,
- programs are producing desired results and achieving intended purposes,
- programs and the legislation authorizing programs should be continued or modified, in the light of experience, to achieve greater economy, efficiency, or effectiveness.

Federal activities designed to evaluate programs have become increasingly significant due to greater demand for program oversight and accountability. This document outlines--from the General Accounting Office's perspective--the status of, and some of the major issues in, Federal program evaluation. We hope that the outlined information and concepts will aid the reader in better understanding the significance, prospects, and problems of Federal program evaluation. We also intend for this document to contribute to the dialogue--in the Congress, Federal agencies, State and local governments, and program evaluation community--concerning how the practice and performance of program evaluation might be improved.

We invite your comments on this document, and would welcome your suggestions concerning how program evaluation might be improved. We hope to use your comments and suggestions in developing future publications on Federal evaluation policy and methodology for dissemination to the Congress, Federal agencies, State and local governments, and the program evaluation community.


Comptroller General
of the United States

October 1978

C O N T E N T S

	page
FOREWORD	
CHAPTER	
1 STATUS OF FEDERAL PROGRAM EVALUATION	1
Interest in Oversight and Accountability	1
Definition of Program Evaluation	4
Aspects of Program Evaluation	5
Scope and Resources for Federal Evaluation	7
2 OVERVIEW OF ISSUES INVOLVING FEDERAL PROGRAM EVALUATION	11
A. Issues in Federal Evaluation Management and Policy	12
A-1. The Lack of a Federal Evaluation Policy: Definitions, Objectives, and Criteria	12
A-2. Incentives and Accountability for Program Performance and Results	14
A-3. Design and Organization of the Evaluation Function	16
A-4. Evaluation Planning and Management	18
B. Issues in the Conduct of Evaluation Studies	19
B-1. Guidelines for Evaluation Studies	19
B-2. Professional Standards for Evaluators	21
B-3. Examining Program Implementation and Process	23
B-4. Using Appropriate Evaluation Criteria	25
B-5. Applying Appropriate Research Methodologies	27
B-6. Adequacy of the State-of-the-Art of Evaluation Methodology	30
3 FUTURE GAO LINES OF EFFORT IN EVALUATION METHODOLOGY AND GUIDELINES	33
FOOTNOTES	35

CHAPTER 1

STATUS OF FEDERAL PROGRAM EVALUATION

Several trends in the 1970s have increased interest in program oversight and accountability. The United States Congress has taken steps to improve its oversight of Federal programs, and is currently considering proposals for further improvements in program oversight and accountability. As a result, various information gathering and analytical activities, sometimes referred to as "program evaluation," have become increasingly significant. In this chapter we will discuss each of these areas, as well as existing estimates of the scope and resources for Federal program evaluation. A definition of program evaluation will also be proposed.

Interest in Oversight and Accountability

Political leaders, public administrators, and the electorate need information with which to oversee programs if our system of government is to operate as intended. A fundamental tenet of a democratic society holds that governments and agencies entrusted with public resources and the authority for applying them have a responsibility to render a full accounting of their activities. This accounting is necessary to enable elected representatives to supervise and control administrative action; to enable administrative officials to effectively manage the programs entrusted to them; and ultimately, to enable citizens to determine the effectiveness with which they have been served by their government.

Several trends in the 1970s have increased interest in Federal program oversight and accountability. More apparent trends include

- reforms in the Federal budget process, such as enactment of the Congressional Budget Act, which established new congressional procedures for budgetmaking; another change was the institution of zero-base budgeting in the executive branch;
- growing concern over the size, scope, and complexity of the Federal Government, the perceived uncontrollable nature of the Federal budget, and the diminishing portion of the budget available for new programs; and

--rising dissatisfaction with apparent government waste, inefficiency, and ineffectiveness manifested in the growing support for tax limitation and "sunset" initiatives.

In the 1970s the Congress has taken steps to improve its capability to oversee the Federal Government. Besides implementing new procedures for budgetmaking, the Congress has increased its staff support for Members and committees, created two new congressional support agencies (the Congressional Budget Office and Office of Technology Assessment), and expanded the analytical functions of the Congressional Research Service and our office.

In many pieces of legislation in the 1970s the Congress has included requirements for programs to be periodically reauthorized and for agencies to study and report to the Congress on the performance of programs authorized by the legislation. In 1978 the Congress enacted legislation establishing an Office of Inspector General in various Federal departments and agencies to review and report on programs.

During the 1970s the Congress has also expanded the oversight responsibilities and authorities of its standing committees. In 1970, committees were required to report on their oversight activities during each Congress. In 1974, committees were authorized to carry out their oversight by contract or by requiring a government agency to do so. Each House has also adopted new rules giving committees additional oversight responsibilities.

Various legislative proposals for further improving oversight and accountability of Federal programs were considered in the 95th Congress. Perhaps the most comprehensive oversight reform proposal considered was S.2, the Sunset Act of 1978, which passed the Senate on October 11, 1978.

The July 13, 1978, report on S.2 by the Senate Committee on Rules and Administration (Senate Report 95-981) states that

"The purpose of the bill, as reported by the Committee on Rules and Administration, is to improve the effectiveness and efficiency of the Federal Government by strengthening congressional procedures for the review and reauthorization of Federal programs. These purposes are to be achieved through the establishment of:

- (1) a 10-year schedule over which all programs, with certain specific exemptions, are subject to termination unless reauthorized following a program review;
- (2) an inventory of Federal programs, updated at the end of each session of Congress;
- (3) procedures for the selection of specific programs each Congress for indepth evaluation; and
- (4) the creation of a 3-year study commission to recommend to the Congress and to the President improvements in Government operations and organizations.

"The committee believes the legislation it has reported is responsive to the clear public demand for improvement in the effectiveness and efficiency of Government programs. The widely heralded "tax-payers revolt," including most notably the voter approval in California of a tax limitation referendum, is open to varied interpretations, but at a minimum, it is clear that taxpayers expect more rigorous, critical, and systematic evaluation of Government spending programs.

"The committee believes the legislation it has reported will meet these objectives and provide the Congress with an improved ability to identify and eliminate obsolete or ineffective programs, reduce wasteful spending, and improve the effectiveness and efficiency of Government programs generally."

Various information gathering and analytical activities in the Federal Government, sometimes referred to as "program evaluation," have become increasingly significant due to the growing demand for better program oversight and accountability. To a large extent, improvements in oversight and accountability will depend on the ability of program evaluation to produce the information and analysis required.

Definition of Program Evaluation

It is difficult to define the status of Federal program evaluation because no universally accepted definition of program evaluation exists. Terms such as "auditing," "program results auditing," "performance auditing," "implementation review," "evaluation research," "program monitoring," "program budgeting," "program analysis," "policy analysis," "policy research," "systems analysis," "cost-benefit analysis," "program review," and "social experiments" have been used synonymously with the term "program evaluation." This situation prompts acceptance of a broad definition.

The GAO, in classifying its own study projects, defines program evaluation as studies of programs which are effectiveness-oriented. They address and are mainly concerned with outcomes--what has been, is being, or should be accomplished through existing Federal programs and activities in relation to objectives established by Congress through statute or by agencies through implementing regulations or procedures. In GAO evaluations, the staff seeks to demonstrate that a program either is or is not accomplishing what it's supposed to be accomplishing. Obviously any evaluation of existing programs has (or should have) future policy implications--but the main objective of GAO program evaluation studies is to examine the operation and outcomes of ongoing programs.

Because stated program objectives--contained in legislation and executive policy documents and subsequently translated in agency regulations and procedures--may be unclear or may not relate to all pertinent program outcomes, an evaluator may have to use other appropriate criteria. In cases where a program produces significant outcomes that were not anticipated when the program was initiated, the evaluator might find it appropriate to identify these unanticipated outcomes. The views, beliefs, and expectations (ascertained through surveys, interviews, or correspondence) of congressional committees, agency officials, program managers, interest groups, and the public may need to be used as evaluative criteria when stated objectives are determined to be inadequate for evaluation.

For these reasons, it should be useful to consider program evaluation as the process of appraising the manner and extent to which programs are

- achieving their stated objectives;
- meeting the performance perceptions and expectations of responsible public officials, interested groups, and/or the public; and

-- producing other significant effects of either a desirable or undesirable character;

to assist future policy and management decisions.*

Three points should be noted about the definition of program evaluation. First, evaluation is a retrospective process; it is always concerned with measuring the performance of an existing program. Evaluation does not necessarily include a prospective element; the evaluation may, but need not, be concerned with identifying and assessing alternative proposals. Second, evaluation does not assign values to a given program; the hypotheses and criteria for evaluation must come from sources other than the evaluator. In an evaluation, a program and its outcomes are tested against objectives, expectations, or values assigned by others (e.g., responsible legislative, executive, or judicial officials in Federal, State, or local governments, individuals, and groups affected by the program). Third, evaluation is a feedback loop in the continuing process of policy and program development, execution, refinement, and reconsideration. In this sense, evaluation is use-oriented; evaluation fails when it does not assist future policy and management decisions.

Aspects of Program Evaluation

The lack of a precise definition of program evaluation should not obscure the actual methods involved in studying programs. There can be many different emphases in evaluation studies. The selection of the appropriate evaluation emphasis and the corresponding techniques and methodologies should depend upon the purpose of the evaluation, particular questions and issues involved in the program, and the stage of the program's implementation.

Notwithstanding the many emphases in evaluation studies, the following three general aspects of program evaluation should be useful.

*This definition is based on the definitions developed by the Office of Management and Budget (OMB) and by program evaluators at the Mitre Corporation symposium on the "Use of Evaluation by Federal Agencies." See "Resources for Federal Program Evaluation: Fiscal Year 1977," Office of Management and Budget, 1977, and Chelmisky, "An Analysis of the Proceedings of a Symposium on the Use of Evaluation by Federal Agencies," Symposium Report, Volume II, M77-39, Metrek Division, Mitre Corporation, July 1977.

- I. Examination of Program Implementation/Process: Examining whether the implementation and execution of actual program activities and operations (processes) meet the perceptions and expectations of responsible political officials and individuals and groups affected by the program, and are in compliance with applicable laws, regulations, and guidelines governing the implementation and operation of the program.

Evaluators should be concerned with program implementation and processes. There is no point in being concerned with the outcomes of a particular program unless one knows that it did indeed take place. How a program is actually carried out is also important to understand. For a variety of reasons, a program in actual operation may not match the program as originally authorized. Actual program operations may not even be designed or able to produce the outcomes deemed desirable and feasible when the program was authorized.

- II. Appraisal of Program Outcome: Examining whether programs have outcomes which achieve legislative/program objectives, meet the perceptions and expectations of responsible political officials and individuals and groups affected by the program, and/or produce other significant effects, whether desirable or undesirable. An evaluation of program outcome may include one or both of the following types of studies:

1. Primary Results Review: Examining whether program activities and operations have immediate, primary, and direct results. Immediate and direct results generally relate to management goals for the program; program performance on such goals can usually be controlled by management.
2. Long Run, Secondary, or Indirect Impact Review: Examining whether program activities and their immediate and direct results are producing or contributing to impacts and effects which achieve long term, secondary, or indirect objectives, perceptions, and expectations, or which are considered desirable or undesirable, by responsible political officials and individuals and groups affected by the program.

Outcome evaluations gauge the extent to which a program effects changes in desirable (or undesirable) directions. Results or impacts are identified, measured, and compared with objectives, desired accomplishments, or expected results.

We have divided outcomes into two classes -- primary results and long term, secondary, or indirect impacts. This conveys the notion that a program, while effective in producing desired primary results, may not be effective in producing intended long term, secondary, or indirect impacts and may also produce undesirable long term, secondary, or indirect effects. Primary results usually relate to agency management goals for a program, while long term, indirect, and secondary impacts usually relate to ultimate values and objectives such as reducing crime, inflation, or unemployment, enhancing health status, maintaining national security, or increasing the economic self-sufficiency of disadvantaged groups. A theory or model linking program activities, short term results, and long term, secondary, or indirect impacts must be identified or developed and then tested to conduct evaluations of long term, secondary, or indirect impact.

III. Assessment of the Relative Effectiveness of Alternatives: Identifying and examining two or more programs or program strategies to determine which is more effective. This type of evaluation may compare program processes, costs, results, and/or impacts, or may compare the relative effectiveness of alternative proposals for changing the program or for new programs.

Evaluation of alternatives usually has a futuristic orientation and is concerned directly with the identification and assessment of alternatives for solving current or potential "problems" and/or improving Government efficiency or performance in the future whether or not a program currently exists. Evaluations of alternatives often are a component of program evaluation studies. In fact, whenever a suggestion or recommendation is made on the basis of an evaluation, the evaluator has (at least implicitly) identified a preferred alternative for improving future program efficiency or performance.

Scope and Resources for Federal Evaluation

The absence of an agreed-upon definition of program evaluation has precluded development of agreed-upon estimates of the number of Federal agency evaluation programs,

offices, or units, the amount spent on the evaluation of Federal programs, the number of evaluation studies, and the number of Federal evaluators. For example, the Office of Management and Budget (OMB) and National Research Council (NRC) surveys of evaluation spending in the Federal executive agencies (discussed below) differed considerably - OMB estimated \$243 million in fiscal year 1977; NRC estimated \$62 million in fiscal year 1976. Both surveys used differing, but narrow, definitions of program evaluation.

The OMB survey of resources devoted to program evaluation by executive branch agencies in fiscal year 1977 found that over \$243 million were obligated by the executive branch for program evaluation, and that over \$169 million, or 70 percent of these funds, were for contracts or grants.

OMB Estimate of Obligations for Program
Evaluation in Fiscal Year 1977
(\$ millions)

	Contracts and Grants	Salaries and Personnel Benefits*	Other Expenses	Total
11 Cabinet Departments	138.7	37.4	8.3	\$ 184.4
28 Major Independent Agencies	31.1	23.8	3.8	58.7
TOTAL	169.8	61.2	12.0	\$ 243.0

Lines may not add to totals because of rounding.

* Does not include evaluation activity by part-time or temporary employees, or those who spend 50 percent or less of their time on program evaluation.

The considerable Federal expenditures for contracts and grants--over \$169 million--has supported a program evaluation industry comprised of a number of organizations such as universities, "think tank" institutions, private firms, and individual consultants.

The OMB survey also showed that over 2,100 executive agency staff years were devoted to evaluation activities (see next page). Virtually every major department and independent agency conducted what OMB defined as "program evaluation."

OMB Estimate of Agency Personnel Resources
Devoted to Program Evaluation in
Fiscal Year 1977

	Agency Full Time Permanent (FTP) Staff Years for Program Evaluation	Evaluators As a Percent of Total FTP Personnel in Agencies	Average Cost (salary plus benefits) of FTP Staff Devoted to Program Evaluation
11 Cabinet Departments	1,422	.09	\$ 26,300
28 Major Independent Agencies	754	.21	31,600
TOTALS	2,176	.11	\$ 28,100

NRC found that over \$1.8 billion was spent on social knowledge production and application activities in fiscal year 1976 of which \$62 million was for "program evaluation" activities (see below).

NRC Estimate of Federal Social Knowledge
Production and Application Activities
Fiscal Year 1976 (\$ millions)

Activity	\$	%
Knowledge production		
Research	655	36
Demonstrations for policy formulation	204	11
Program evaluation	62	3
General purpose statistics	294	16
Total	1,215	67
Knowledge application		
Demonstrations for policy implementation	183	10
Development of materials	121	7
Dissemination	294	16
Total	598	33
TOTAL	1,813	100

Lines may not add to totals because of rounding.

The OMB and NRC surveys did not include the Department of Defense's extensive technical evaluations of major weapons and support systems. Evaluation activities in the legislative branch agencies, including our office, were also not surveyed.

In considering the scope and resources for Federal evaluation, acknowledgement should be given to the resources expended for internal audit activities within Federal agencies. Although much of the work of agency internal audit groups may not be program evaluation (e.g., financial audits), omitting all of the internal audit resources in the Federal Government would result in an incomplete picture of program evaluation activities within Federal agencies. Over 11,600 personnel are devoted to internal audit functions within Federal agencies, but little research has been done on the interrelationship of internal audit and program evaluation.

Similarly, recognition should also be given to State and local audit and evaluation resources devoted to the examination of Federally assisted programs in those jurisdictions. Most States and many local jurisdictions have established units that review and analyze the performance of government programs, and over one-half of the State legislatures have enacted some form of "sunset" legislation. State and local program review units range from traditional post-audit agencies to multiple purpose agencies that carry out various types of research on programs.

CHAPTER 2

OVERVIEW OF ISSUES

INVOLVING FEDERAL PROGRAM EVALUATION

Trends in the 1970s favoring oversight and accountability of Federal programs also favor oversight of Federal evaluation programs, activities, and expenditures. This chapter outlines a variety of issues likely to arise in the oversight of Federal program evaluation. Such issues will also need to be resolved if the performance of program evaluation is to be improved. Issues in program evaluation arise in the context of describing, appraising, and identifying problems and needed improvements in the practice of evaluation.

The practice of Federal program evaluation can be viewed as involving two basic aspects. These include

- practices related to the organization and management of evaluation programs or offices in the Federal government, and
- practices related to the conduct of individual evaluation studies.

The issues involving Federal program evaluation can thus be classified into two major groups:

- A. Issues in Federal evaluation management and policy. These are concerned with the role, objectives, capabilities, and organization of evaluation in the Federal Government and the effectiveness of processes by which Federal agencies plan and manage their evaluation programs.
- B. Issues in the conduct of program evaluation studies. These are concerned with the effectiveness of methods used to carry out individual evaluation studies of Federal programs and the adequacy of the state-of-the-art of program evaluation.

Individual issues within these two major groups will be discussed separately in the following section. To support our treatment of the issues, we have attempted to use material, including direct quotes, from literature expressing the concerns of a wide sampling of the Federal evaluation community.

A. ISSUES IN FEDERAL EVALUATION MANAGEMENT AND POLICY

A-1. The Lack of a Federal Evaluation Policy: Definitions, Objectives, and Criteria

It is difficult to satisfactorily measure, appraise, or improve the effectiveness of current Federal evaluation management and policy. A major reason for this is the lack of consensus in the Federal evaluation community (Congress, GAO, OMB, Federal agencies, and non-Federal evaluation producers and users) concerning an operational Federal evaluation policy. This includes agreed-upon definitions, objectives, and criteria for evaluation, and in particular, the roles and objectives of individual evaluation groups in the legislative and executive branches of the Federal Government. Agreement on the types of evidence required to demonstrate the success of evaluation groups, an integral component of any Federal policy for evaluation, is also lacking.

The lack of a Federal evaluation policy has, according to some experts, led to uncertainty and conflict on the part of evaluation offices as to how to organize, plan, and manage their work.^{1/} Evaluation offices in the Federal Government--including GAO--can experience frustration by conflicting demands with no obvious way to choose between them. The lack of agreement on the measures and standards for appraising evaluation offices among those who plan, manage, conduct, and review evaluation efforts in the Federal government has caused and will continue to cause problems.

There are many viewpoints on the appropriate role of evaluation in the Federal Government. Chelimsky has classified the different standpoints into three broad perspectives:

"A knowledge perspective: In this view, the purpose of evaluation is to establish evidence leading to new knowledge about...problems and about the effectiveness of governmental strategies for addressing them....

"A management perspective: In this view, the purpose of evaluation is to serve as a flexible management tool, a support system for assessing and improving the operational efficiency of government programs....

"An accountability perspective: In this view, the purpose of evaluation is to make the best possible use of tax resources by holding program managers accountable for the worth (i.e., both the effectiveness and operational quality) of their programs."^{2/}

An evaluation office or an evaluation study of a Federal program may serve many audiences including

- Congress (including congressional committees, individual members and staff);
- executive branch policymakers (including the President, the Cabinet, and agency heads and their staff, e.g., OMB, NSC);
- Federal agency policy officials, program managers, and their staff;
- State, local, international policy officials, program managers, and their staff;
- individuals and groups directly served, regulated, or affected by a program; and
- public interest groups, the press, the general public, and other evaluators.

Although the information needs of users in these groups might overlap, in reality the differing functions and roles of the potential audiences for an evaluation tend to generate highly disparate expectations about what information an evaluation should produce.^{3/} Federal "program evaluation" clearly signifies different things to different audiences.

All of the information requirements implicit in the different perspectives and audiences cannot be met simultaneously by an individual evaluation study or office. Consequently, evaluation studies and offices are and will continue to be subject to criticism from those who hold different perspectives on the role of evaluation and on the extent to which the needs of the various potential audiences for an evaluation should be served.

The lack of an agreed-upon operational Federal evaluation policy in part results from these differing perspectives and from a lack of understanding and agreement on other evaluation policy and management issues that have been raised in the evaluation community. These issues will be discussed in the following sections. Because of these differing perspectives, there is not a consensus within the evaluation community on these issues. However, we believe these issues represent the major concerns that have been expressed concerning evaluation policy and management.

A-2. Incentives and Accountability for Program Performance and Results

At the November 1976 Mitre Corporation symposium on the use of evaluation by Federal agencies, the consensus among Federal agency personnel was that there are few incentives within agencies they represented (and in the Federal Government generally) to encourage program effectiveness through the accountability of program managers.^{4/} Agency representatives indicated that, presently, program managers are not held accountable for program performance. One agency representative stated:

"Need I tell you that it often seems that there is accountability for everything but results? There is accountability for fidelity to a policy line even when the policy is vague or ill-defined. There is accountability for good public and congressional relations. There is accountability for spending one's money promptly. There is accountability for assuring compliance with a thousand and one Federal laws and regulations and so on. But to make Government work, we must establish accountability of Government managers for program performance and program results."^{5/}

In reviewing government programs, evaluators have often seen one of the consequences of the lack of incentives and accountability--programs not being managed to achieve results on certain objectives. According to Scanlon, this situation is characterized when evaluators find

- different and conflicting viewpoints on what the program objectives and activities are,
- unexpected gaps in program operations and implementation,
- priority congressional or agency objectives receiving no program design attention,
- no capability for receiving or using evaluative information to maintain or improve program performance, and/or
- an absence of formal and informal channels providing feedback on stated objectives.^{6/}

Some observers believe that if program managers were accountable and rewarded for proven program results and performance, they would have an incentive to manage well,

institutionalize the evaluation function, and use evaluation information to improve and document the results and performance of their programs.^{7/} Presently, evaluation information will not be used unless a decisionmaker wants to use it, and it happens to be available. Without incentives and accountability for program results, decisionmakers often will have no desire to request and little enthusiasm for using evaluations. According to Scanlon, in the present environment, the evaluator's

"...chances of improving government policies, programs and productivity by making evaluation information available are extremely low (and)...attempts to improve evaluation practices and methods are working a dry hole."^{8/}

The establishment of a Senior Executive Service for Federal "super-grade" employees represents one approach for strengthening management incentives and the accountability of program managers.

A-3. Design and Organization of the Evaluation Function

A variety of questions have been raised in the evaluation community concerning the proper design and organization of the evaluation function in the Federal Government. These questions relate to the general design and organization of evaluation responsibilities both in the legislative and executive branches. These questions have been raised in the consideration of oversight reform proposals, including proposals to authorize or require evaluation in individual agencies and for individual programs. Some of the major questions include

- audiences to be served by evaluation--To what extent should the evaluation function be organized and designed to serve different potential users of evaluation?
- organizational structure of the evaluation function--To what extent should evaluation be centralized or decentralized in an agency, in the executive or legislative branch, or in the Congress itself?
- independence of the evaluation function--To what extent should the evaluation function be independent of congressional committees, executive policy officials, or program managers?
- mix of in-house vs. contract evaluation work--To what extent should evaluation studies be done by outside contractors?
- choice of group or agency responsible for evaluating a particular program--Should the agency responsible for administering the program, a related program agency, a research agency, an advisory committee, a special commission, OMB, GAO, or a congressional committee have the lead responsibility for evaluation?
- the funding of evaluation--How much should be spent on evaluation, should funds be earmarked for evaluation, and should a certain percentage of program funds be set aside for evaluation?
- guidelines and authority for the evaluation function--Should OMB or the Congress establish explicit requirements for the evaluation of particular programs or for the evaluation function in agencies?

--privacy of data on individuals and confidentiality of the evaluation function--To what extent should information on the operation of the evaluation function, evaluation data, and evaluation results be made available to various interested groups outside of the agency responsible for evaluation?

How one answers questions related to the design and organization of the evaluation function depends on one's viewpoint on the proper role and objectives of evaluation (e.g., knowledge vs. management vs. accountability viewpoints). Those who see program evaluation from a knowledge or accountability viewpoint, for example, would tend to favor a relatively independent, non-confidential evaluation function so that objective and technically competent evaluations could be conceived, carried out, and reported to advance knowledge or enable accountability. However, if an evaluation office tried to be totally independent irrespective of the interests (e.g., reputations) of program or policy officials, the evaluation office might meet resistance from officials or managers responsible for the programs being evaluated, might become separated from the "real" decision processes, and might come under pressure to conduct studies of only minor non-controversial programs or activities so that the evaluations do not jeopardize the position of officials or managers.

On the other hand, persons viewing program evaluation from a management perspective would tend to favor a relatively confidential evaluation function closely linked to policymakers and line managers. However, if an evaluation office is totally subservient to policymakers or program management, evaluation results which make policy officials or program management "look bad" might be suppressed, high quality studies might have a low priority in comparison to "fire-fighting" projects, and it might be difficult for others to appraise the evaluation operation because of a lack of documentation and the confidentiality of evaluation operations and results.

A-4. Evaluation Planning and Management

A variety of questions have been raised in the evaluation community concerning the adequacy and effectiveness of evaluation planning and management. These questions relate to the processes by which an evaluation function or office assures that it achieves its objectives. Evaluation planning and management includes processes used to

- set evaluation office objectives and evaluation priorities;
- select programs and topics for evaluation;
- assure proper coordination and minimize unnecessary duplication with other evaluation groups;
- allocate resources and staff;
- acquire evaluation information by contract;
- monitor evaluation efforts to assure acceptable quality, usefulness, and resource expenditures;
- package evaluative information;
- disseminate evaluation information; and
- appraise results and performance (usefulness, timeliness, quality) and lessons learned.

Concerns have been expressed with the adequacy of these processes. For example, Abt has been critical of the processes used to set evaluation priorities and allocate evaluation resources ("the highest cost programs do not get proportionally evaluated now; we need to allocate in proportion to impact"), processes used to select issues for evaluation ("evaluations...rarely compare programs with similar goals in different departments"), processes used to acquire evaluation information by contract ("there are unclear statements of specifications for the operators of program evaluation...work statements of program evaluation requests for proposals describe what they want done and how but not why"), and processes used to disseminate evaluation information ("there is no significant...planned connection between outputs of program evaluations and the major programmatic budgetary decisions about programs").^{9/}

B. ISSUES IN THE CONDUCT OF EVALUATION STUDIES

B-1. Guidelines for Evaluation Studies

Questions concerning the need for and appropriateness of evaluation study guidelines have been raised in the evaluation community. While there is a consensus on the need to improve and transfer state-of-the-art evaluation methodology to enable more cost-effective--efficient, timely, credible, useful--evaluations to be produced, there is not complete agreement on guidelines as the appropriate means for meeting this need.

On the one hand, it is argued that guidelines are needed to

- improve the efficiency, timeliness, and consistency of practice;
- help avoid misapplications of the state-of-the-art;
- assure the acceptability, credibility, and quality of evaluation information provided decisionmakers; and
- enable judgments to be made of the work of evaluators.

Users of evaluations should be able to assume that evaluations will be performed competently in a timely manner, and will be reported accurately. In line with this reasoning, we have developed general guidance for evaluation studies. The GAO report, Evaluation and Analysis to Support Decisionmaking (PAD-76-9), September 1, 1976, discusses certain conceptual aspects of conducting evaluations, as well as various practical aspects of managing and performing evaluation studies. Finding Out How Programs are Working: Suggestions For Congressional Oversight (PAD-78-3), November 22, 1977, outlines a process for the sequential review of the separate stages of an agency's implementation of legislation (e.g., executive policy, program design, establishment, operation, results, and impacts) to enable the identification and reporting of any major problems in a program or deviations from legislative intent before large amounts are invested in the program or in data collection efforts.

On the other hand, concerns have been raised regarding the appropriateness of guidelines. No one disputes the fact that the quality of program evaluation has improved in the past decade, in spite of the scarcity of generally

accepted guidelines for evaluators. Other concerns--typical for the guidelines issue in any context--must be recognized and accounted for if and when guidelines are developed, including concerns that guidelines for evaluation studies may

- discourage experimentation and innovation needed for developing improvements in evaluation methodology,
- constrain evaluators from tailoring studies to particular situations,
- unnecessarily restrict the supply of certain types of evaluation and drive up the costs of producing evaluation information,
- not be comprehensive enough to cover all the circumstances to which they will be applied,
- be difficult to enforce because their scope makes it impossible to monitor compliance,
- require constant updating to keep up with advances in the state-of-the-art, and
- be counterproductive if they deflect efforts from the achievement of study objectives to compliance with the guidelines.

The absence of a consensus on the need for evaluation study guidelines results, in large part, from the absence of an agreed-upon Federal evaluation policy. Differing viewpoints on the role of evaluators has resulted in wide variations in the evaluation methods used to develop evaluative information and the resultant quality of such information.

Essentially, these differing viewpoints require evaluators to make tradeoffs--between the study objectives of timeliness, cost, validity, reliability, scope, comprehensiveness, pertinence, and significance--in the conduct of particular studies. Depending on the situation, acceptable evaluations can range from a 1 day site visit (or a few telephone calls) to a multi-year multimillion dollar longitudinal social experiment. In other words, the methodologies which should be employed in a study depend on the purposes of the study and the use to be made of the study results.

B-2. Professional Standards for Evaluators

Questions concerning the need for and appropriateness of professional standards for evaluators have also been raised in the evaluation community. As noted previously, there is a consensus on the need to improve and transfer state-of-the-art of evaluation methodology to enable more cost-effective evaluations to be produced. However, there is not a consensus on standards as the appropriate means for meeting this need.

On the one hand, proponents of standards argue that users of evaluations should be able to assume that evaluations were performed competently and reported accurately. Assumptions about quality presently rely principally upon the academic credentials of the performers and/or the reputation of the organizations they represent. More than that is needed. If evaluations are to be accorded credibility and receptivity as professional work, there should be professional standards behind them. Professions such as law, medicine, and accounting have found that academic credentials and individual and institutional reputations are not sufficient and had to be supplemented by formal professional standards. Since program evaluation may be interdisciplinary, both with respect to subjects and performers, it is difficult to develop the kind of consensus needed to derive and support such standards. But notwithstanding interdisciplinary interests, the following can be defined

- adequate technical training and proficiency,
- independence in mental attitude,
- due professional care,
- sufficiency and competency of evidential matter,
- consistency of practice, and
- adequacy of disclosure.

These items roughly parallel generally accepted auditing standards and represent the kinds of quality assurance that users of evaluation affecting policy should reasonably expect.

While the art of program evaluation has moved visibly in the direction of professionalization in the last decade or so, proponents of standards argue that it will be to the mutual advantage of evaluators and the users of their products, and especially government users, if there is now movement toward defined standards.

On the other hand, the concerns regarding the appropriateness of guidelines also apply to standards, particularly the concerns that they may unnecessarily restrict the supply of evaluators and increase the costs of producing evaluation information. Other concerns regarding standards have been raised, including concerns that:

- It may be too early in the evolution of the evaluation profession to develop standards.
- Standards could, in effect, represent licensing requirements for evaluators, with all of the attendant problems of monopoly and self-regulation.
- The evaluation field should not be considered in the same sense as the legal, medical, or accounting professions, but more like the profession of journalism.
- Standards relating to the sufficiency of evidential material may limit an evaluator's ability to develop information in time for use in a decision.

Nevertheless, significant efforts are being initiated to develop evaluation standards by such groups as the Evaluation Research Society of America.

B-3. Examining Program Implementation and Process

A general criticism of the present state of evaluation is that evaluators tend to do "black box" evaluations (i.e., evaluations comparing only program inputs, such as dollars appropriated for a program, and program impacts, such as changes in average lifetime earnings of certain groups). Black box evaluations do not confirm how program inputs are translated into actual program activities, processes, or treatments.

Program implementation and process is important to examine when assessing program impacts. Such examinations study the steps that occur between a policy mandate for a program (e.g., authorizing or appropriations legislation) and the occurrence of impacts by asking whether and how

- a program was initiated;
- a program deviated in actual practice from what was initially authorized and planned;
- actual program activities are designed to achieve the impacts originally intended; and
- useful measurements can be made of the relationship between actual program activities and program impacts.

If the program inputs are not translated into the desired activities, and if the activities are not actually carried out as expected, then a simple examination of impact may be disappointing and misleading. Without confirming how a program is carried out at the field or operating level of a program, one cannot--without great risk--conclude that certain impacts occur as a result of the program. According to Freeman:

"Failure to study process is serious. There may be many programs that offer significant solutions to defects in the human condition that have been passed over because the evaluation failed to show an impact. Yet, in fact, the real reason for the lack of impact was that the program was never implemented fully well, or at all."10/

In the case of aid-to-education programs initiated by legislation in the 1960s, Yin notes:

"...the implementation process had frequently changed the nature of the programs and hence the nature of the expected outcomes. New projects varied particularly

in the degree to which they provided teachers with initial training, allowed for adequate staff meetings, and monitored changes in teaching practices...In a few cases, projects changed so much during implementation that the children of low-income families were no longer even the actual beneficiaries."11/

Others have made similar observations. Coleman has called for the conduct of "social audits" which trace the flow of resources for a program down to the level at which program services are delivered to the target groups.12/ The Urban Institute has begun an institutional analysis program to develop more systematic ways of studying program implementation.13/

Besides guarding against faulty conclusions about the impact of a program that was never implemented as intended, examination of program implementation and process would also help reduce the tendency of evaluators to focus on contextual (uncontrollable) variables instead of variables under the control of responsible program managers or policymakers.

B-4. Using Appropriate Evaluation Criteria

One issue often raised in the conduct of evaluation studies concerns criteria that the evaluator should use in evaluating a particular program.

Evaluators often encounter difficulties in attempting to identify criteria for evaluation. This, in part, results from a lack of agreement on program objectives and on the types of information needed to verify program performance. Different committees of the Congress, agency officials, State and local officials, program personnel, interest groups, etc., may have different beliefs about what a program is and should be doing, and what data is sufficient for proving program "success." Evaluators must try to answer evaluation questions and/or try to reach a conclusion regarding program effectiveness when

- legislative intent and stated program goals are vague, appear to conflict with each other, appear to be "symbolic" rather than real, or have not been translated into operational terms by the agency;
- available evaluation criteria and standards appear inconsistent with legislative intent or actual program activities; and
- there appears to be disagreement among the members of Congress, congressional committees, executive agency officials, and/or State and local officials over what the program is intended to accomplish and what standards should be used to define program performance and program "problems."

Federal policies do exist by virtue of statutory language and intent, guidances from the President and agency policy officials, individual program plans, guidelines, and regulations, position papers from the Office of Management and Budget, and chosen evaluation criteria. Often, however, they all differ, resulting in a continuing debate concerning program objectives and success criteria.¹⁴ In other words,

- specific program goals and the evidence required to verify performance in achieving such goals are seldom articulated and agreed to by relevant decisionmakers (e.g., Congress, executive policymakers and program managers);

--the logic linking program inputs (resources), processes, outcomes, impacts, and goals is seldom articulated and agreed to; and

--sufficient decision making authority (power to acquire and allocate resources...) is seldom vested in program managers.15/

The evaluator is one of the first people (and sometimes the only person) who must produce, through an actual measurement of concrete "real world" situations, answers to questions formed from more rhetorical statements in the political world (e.g., higher levels in the agency or in the Congress). The evaluator must determine from the rhetoric of those in charge what was the legislative and policy intent, and then must make measurements to see if the actual program addresses this intent. In this environment, Abt notes that often evaluators make the following mistakes:

1. Mis-selection of program goals and subsequent criteria of effectiveness, as in cases where the program operator's goals differ from those of the sponsors and possibly those identified by the evaluators.
2. Failure to focus evaluations on controllable policy variables (variables under the control of policy makers) instead of contextual variables (variables not under the control of policy makers).
3. Failure to select the readily measurable measures of effectiveness relevant to the policy issues.16/

B-5. Applying Appropriate Research Methodologies

Concerns have been raised in the evaluation community about the extent to which evaluation methodologies are applied properly, particularly in evaluations of program impact. Evaluations of program impact require competent application of research designs which permit inferences to be made about the changes brought about by a program. According to Evans there are two reasons for assuring the competent application of evaluation methodology:

"When we talk about program effectiveness, we are basically talking about a cause-effect question. We want to measure what changes have resulted in connection with the program, but more importantly we want to be able to attribute those to the program, not just the passage of time or some other extraneous variable. That immediately brings you into the matter of research and evaluation design.

"The other reason why design and methodology are so important is because all of the programs that we are talking about (or nearly all of them) are inherently controversial social action type programs. As such, in the political sphere, in the Congress, and in the public they have both their protagonists and their detractors. That means that any evaluation of any of these programs, no matter what it finds--whether it finds the program effective or ineffective--is going to be attacked, not because the findings are distasteful which may be the real reason, but on methodological grounds. Therefore, if the evaluation is not itself methodologically defensible to a reasonable degree, its chances of influencing policies and budgets are thereby lessened substantially."17/

To measure program impacts, many evaluators have tried to replicate the classical experimental model used in scientific research laboratories. Many evaluators believe that the methodological quality of an evaluation is directly related to the degree to which the classic experimental model is replicated. According to Abt, for example, some of the typical weaknesses in applying evaluation methodology include:

- 1--Lack of randomization in the selection of experimental and control groups, biasing the findings.
- 2--Lack of any control groups, rendering the validity and generalizability of the findings completely uncertain.

- 3--Failure to recognize and compensate for statistical artifacts such as regression towards the mean.
- 4--Failure to select the readily measurable measures of effectiveness relevant to the policy issues.
- 5--Too broad a focus, and failure to limit the evaluation to program aspects that can be researched in enough depth to achieve validity within the time and effort resources available - the "fishing expedition" that generates a low signal-to-noise ratio.
- 6--Too narrow focus, with premature concentration on a few variables assumed to be significant at the cost of excluding potentially more significant ones.
- 7--Too short duration to allow sufficient iteration of the data gathering efforts to narrow down to the critical issues in depth.
- 8--Misdesign of questionnaires, often as a result of a lack of pilot testing that may be a result of a lack of sufficient time.
- 9--Data gathering survey misdesign, particularly the inappropriate allocation of survey resources that result in an over-sampling of the most significant groups, as was the case with the attempted application of the Watts-Conlisk sampling model to the New Jersey income maintenance experiment.
- 10--Failure to explore or ignorance of relevant literature and data concerning similar programs or phenomena being evaluated.
- 11--Failure to recognize and design around the Hawthorne effect and other types of self-fulfilling and self-selecting and de-selecting phenomena.
- 12--Voluntarism in the selection of surveyed experimental and control groups leading to biased results, as when unsuccessful programs choose not to be a part of an evaluation.
- 13--Premature and inappropriate data collection efforts of perishable information that cannot subsequently be readily corrected, as in the collection of significant amounts of unanalyzable data on planned variation in the Follow-Through Experiment.

- 14--Confusion of a program inventory for a program evaluation--the substitution of merely descriptive case studies for analysis of program effectiveness and efficiency with respect to goals and resources use.
- 15--Confusion of needs assessment for evaluation, assuming that determination of the relevance of program design to an established set of needs constitutes an adequate evaluation.
- 16--Confusion of program operating standards or program outputs for impacts, as is frequently the case in school evaluations that boast of increased daily attendance that may or may not have any relationship to desired educational impacts such as increased lifetime earnings.
- 17--Contamination of program results by formative evaluation procedures that directly influence program activities as part of the evaluation processes in non-reproducible ways.18/

According to Abt, if these weaknesses were reduced or removed, the usefulness of evaluation and its applicability to legislation would be improved.

B-6. Adequacy Of The State Of The Art Of Evaluation Methodology

Questions have been raised in the evaluation community about the adequacy of the state-of-the-art of evaluation methodology. Some have argued that current evaluation methodology is adequate, since the theoretical work in evaluation was completed 30 years ago. Others have argued, for a variety of reasons, that evaluation methodology based on the rational/experimental model is inadequate and that a new evaluation "paradigm" is needed.

Those who argue that current evaluation methodology is adequate basically view evaluation as a process of applying social science, operations research, and survey research methods to answer questions about program impact. Perhaps the strongest proponent of this viewpoint is Abt, who has stated that

"...most of the statistical concepts, the experimental designs, the instrument design issues, the operational designs of program evaluations were done in the 1940s or earlier--many of them in World War II operations research, in earlier public opinion and marketing research, in survey research in the 1930s developed for commercial purposes....

"A lot of the mistakes and errors we've made in social programs evaluations have resulted in the rediscovery of things that have been known scientifically for 20 years but were simply not known to social programs evaluators....

"The evaluation art has advanced to the point today where we can evaluate anything, we can determine benefit/cost ratios, we can determine relative and absolute cost-effectiveness - what we can't do is always get it used and communicate our results."19/

On the other hand, the adequacy of evaluation research methodologies increasingly has been questioned, both from philosophical and practical viewpoints. Schon, for example, has stated that

"Contrary to mythology, we are largely unable to 'know' in situations of social change, if the criteria of knowledge are those of the rational/experimental model. The constraints on knowing affect not only our ability to gain certainty, or probability, or precise knowledge, but our ability to establish knowledge in the rational/experimental mode at all."20/

And according to Guttentag:

"The context within which evaluation research is conducted is too different from the classical experimental model. Most of the assumptions of the model cannot be fulfilled... the researcher abstracts those few hypotheses which he considers 'researchable' from the program's broad and multifaceted goals. Given the researcher's constraints, the final hypotheses may have little relevance to what is actually going on in the program. In addition, the researcher cannot randomize subjects or treatments. He cannot control the flow of subjects into or out of programs. When he does try to do so, conflicts with program administrators result."21/

"One way in which this issue has been handled by researchers trained in the older (scientific) model, is not to have anything to do with a program unless program administrators promise to make it an independent variable...."

"When it is demanded that problems be forced to fit methods, rather than vice versa, then the need for a paradigm shift is patently clear...."(emphasis added)

"When the older (scientific) paradigm shapes the discussion in the evaluation literature...(such issues) are relegated to the annoying, interfering category of 'the politics of evaluation'--which researchers believe often interrupts what they think to be the proper conduct of evaluative research...."

"We do not agree that evaluation research is a pallid, warmed-over version of applied social research. On the contrary, its scientific, intellectual problems are distinctively new, because of the new contexts, orientations and issues confronted in post-industrial societies...."22/

An increasing number of evaluators now view evaluation as an integral part of the decisionmaking and government learning function of gathering feedback information--whether by experimental or any other methods.23/ Researchers at the Urban Institute, for example, have developed a set of evaluation techniques, designed to be used at different stages of a program's life,

CHAPTER 3

FUTURE GAO LINES OF EFFORT IN

EVALUATION METHODOLOGY AND GUIDELINES

Title VII of the Congressional Budget Act of 1974 implicitly requires our office to exercise a leadership role in Federal program evaluation. Our initial implementation of the Title VII responsibilities has contributed to our understanding of the issues, problems, and concerns in Federal evaluation policy and methodology outlined in the preceding sections. In the long run, however, to efficiently and effectively carry out our Title VII responsibilities, we will need to do more than just identify the status, issues, and prospects of program evaluation. We--in cooperation with the Federal evaluation community--will need to

- improve our ability to evaluate programs and provide other evaluation related assistance to Congress;
- further the state-of-the-art for making high quality, useful program evaluations;
- identify and help bring about improvements in Federal evaluation management and policy; and
- transfer and share knowledge and lessons learned in evaluation so that Congress, Federal agencies, and State and local governments can improve their program evaluation capabilities.

We believe the climate is good for starting to deal effectively with most of the issues, concerns, and problems in evaluation and we hope that the information and concepts presented in this document will be a useful step toward meeting our Title VII responsibilities. We are currently undertaking work to

- identify evaluation practitioner needs in the methodology area which deserve priority attention;
- identify existing evaluation methods;
- determine where and how existing methodologies require further development;
- develop evaluation methods to meet identified needs;

- demonstrate newly developed methods or improvements of existing evaluation methods to assure a credible basis for wide acceptance in the evaluation community;
- transfer evaluation methods throughout GAO and the evaluation community;
- identify the objectives, functions, policies, organization, planning, and management of program evaluation in the Federal government;
- assist in the development of agreed upon policies and criteria for Federal program evaluation; and
- appraise the performance of Federal evaluation activities according to agreed-upon criteria.

We seek your advice, comments, and cooperation in helping us bring about improvements in the policy, management, and methodology of Federal program evaluation. We expect that some improvements in Federal evaluation will require our leadership, and we are contemplating a series of roundtable seminars as a vehicle for developing a consensus Federal evaluation policy.

FOOTNOTES

- 1/ Schmidt, R., et al., The Market For Evaluation Services in the Department of Health, Education and Welfare, The Urban Institute, Contract Report 962-21, May 1975, pp. 5-6.
- 2/ Chelimsky, E., An Analysis of the Proceedings of a Symposium on the Use of Evaluation by Federal Agencies, Symposium Report: Volume II, Metrek division, The Mitre Corporation, M77-39, July 1977, p. 6.
- 3/ Ibid., p. 6.
- 4/ Ibid., p. 38.
- 5/ Ibid., p. 38.
- 6/ Scanlon, J. "Federal Evaluation Policy: The Cart Before the Horse," The Urban Institute, Working Paper 3-0003-4, May 27, 1977, p. 9.
- 7/ Chelimsky, op. cit., p. 38.
- 8/ Scanlon, op. cit., p. 1.
- 9/ Abt, C., "The State of the Art of Program Evaluation," statement at the Legislative Oversight and Program Evaluation Seminar Series, May 19, 1975, Abt Associates, Inc., pp. 5,7,13.
- 10/ The arguments for standards are excerpted from Patterson, P., "Comments on Clark Abt's Paper," attachment to a GAO memorandum, January 4, 1978.
- 11/ Freeman, H., "The Present Status of Evaluation Research," in Evaluation Studies Review Annual, Vol. 2, 1977, Guttentag and Saar, ed., Sage Publications, 1977, p. 27.
- 12/ Yin, R., "Production Efficiency Versus Bureaucratic Self Interest: Two Innovative Processes?," Policy Sciences, 8 (1977), p. 382.
- 13/ Coleman, J. "Problems of Conceptualization and Measurement in Studying Policy Impact," in Public Policy Evaluation, Dolbeare. ed., Sage Publications, 1975, p. 36-39.

- 14/ Nay, J., et al., Introduction to Testimony before the Senate Committee on Human Resources, The Urban Institute, October 6, 1977, p. 4 and Schmidt, op. cit., p. 7.
- 15/ Horst, P. et al., "Program Management and the Federal Evaluator," Public Administration Review, Volume 34, No. 6, July/August 1974, p. 300-308.
- 16/ Abt, op. cit., pp. 14, 18, 20.
- 17/ Chelimsky, E., Proceedings of A Symposium on the Use of Evaluation by Federal Agencies, Symposium Report: Volume I, Metrek division, Mitre Corporation, M77-39, March 1977, p. 39.
- 18/ Abt, op. cit., pp. 17-20.
- 19/ Ibid, p. 10-12.
- 20/ Schon, D., Beyond The Stable State, W.W. Norton Company, 1971, p. 201.
- 21/ Guttentag, M., "Special Characteristics of Social Intervention Programs: Evaluation of Social Intervention Programs," Annals of the New York Academy of Sciences, Volume 218, June 22, 1973, p. 6, cited in Chelimsky, Vol. II op. cit.
- 22/ Guttentag, M., "Evaluation and Society," in Evaluation Studies Review Annual, Vol. 2, 1977, Guttentag and Saar, ed., Sage Publications, 1977, pp. 58, 59.
- 23/ Ibid., Schon, op. cit., Nay, J., et. al., "Purposeful Behavior Revisited: Or Why Some Measures and Comparisons Are More Important to Some People Than to Others," The Urban Institute, Working Paper 783-12, March 1977 and Argyris, C., "Double Loop Learning in Organizations," Harvard Business Review, Volume 55, No. 5, Sept./Oct. 1977.
- 24/ Urban Institute Annual Report 1977.

Re: Federal Program Evaluation: Status and Issues

Request: Your responses to the following questions will be appreciated.

**Harry S. Havens, Director
Program Analysis Division**

1. How would you rate the general value of the information and issues presented in this report?

- Excellent Above Average Average Poor

2. To what extent were the following sections of this document useful to you?

	<u>Highly useful</u>	<u>Somewhat useful</u>	<u>Not useful</u>
a) Interest in oversight and accountability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Definition/aspects of program evaluation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Scope/resources for Federal evaluation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Issues in evaluation management and policy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Issues in the conduct of evaluation studies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f) Future GAO lines of effort in evaluation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. What topics in this document need more emphasis?

4. What do you view as the major problem in Federal program evaluation?

5. What suggestions do you have for improving Federal program evaluation?

6. Other comments?

Please fill in your return address (opposite) if you would like to receive future GAO publications on program evaluation topics.

Name _____ Title _____
Organization _____
Address _____
City _____ State & Zip Code _____

POSTAGE AND FEES PAID
U.S. GENERAL ACCOUNTING OFFICE



THIRD CLASS

MR. HARRY S. HAVENS
US GENERAL ACCOUNTING OFFICE
441 G STREET NW ROOM 5001
WASHINGTON, DC 20548

Single copies of GAO reports are available free of charge. Requests (except by Members of Congress) for additional quantities should be accompanied by payment of \$1.00 per copy.

Requests for single copies (without charge) should be sent to:

U.S. General Accounting Office
Distribution Section, Room 1518
441 G Street, NW.
Washington, DC 20548

Requests for multiple copies should be sent with checks or money orders to:

U.S. General Accounting Office
Distribution Section
P.O. Box 1020
Washington, DC 20013

Checks or money orders should be made payable to the U.S. General Accounting Office. NOTE: Stamps or Superintendent of Documents coupons will not be accepted.

PLEASE DO NOT SEND CASH

To expedite filling your order, use the report number and date in the lower right corner of the front cover.

GAO reports are now available on microfiche. If such copies will meet your needs, be sure to specify that you want microfiche copies.

AN EQUAL OPPORTUNITY EMPLOYER
U. S. GENERAL ACCOUNTING OFFICE
WASHINGTON, D. C. 20548

OFFICIAL BUSINESS

PENALTY FOR PRIVATE USE, \$300



POSTAGE AND FEES PAID
GENERAL ACCOUNTING OFFICE

THIRD CLASS