

GAO

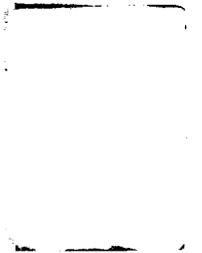
Transition Series

November 1988

Program Evaluation Issues



GAO/OCG-89-8TR



•

Comptroller General
of the United States

B-158195

November 1988

The President of the Senate
The Speaker of the House of Representatives
The Director-designate, Office of Management and
Budget

This summary report is one of a series that addresses major policy, management, or program issues facing the new Congress and administration. The discussion of issues, the problems associated with each, and recommended actions are based on our work in the program evaluation area.

In this report, we are concerned with the government's continuing ability to develop, disseminate, and use sound information. Production of sound and timely information is one of the most critical functions of government. Program evaluation—along with supporting data collection—is one of the best means available for obtaining it. Yet with few exceptions, we have found that both program evaluation and data collection capabilities have been gravely eroded in the executive branch.

Officials in both executive and legislative branches need quality evaluation to help them reach sound judgments. Without this capability, executive branch policymakers are in a weak position to pursue their policy objectives with the Congress, to justify continuation of their programs, and to eliminate wasteful or unnecessary initiatives, because they lack supporting data.

The legislative branch continues to need program evaluation findings, whether or not the executive branch produces them. If the current drawdown in evaluation capabilities continues, reports from the GAO, its sister agencies, and from private sector analysts may become the only sources to which the Congress can turn for sound information to guide key decisions. The erosion of evaluation capability

in the executive branch will not insulate agency programs from congressional oversight. On the contrary, lacking their own evaluations, agencies could find themselves excluded from meaningful participation in congressional decisions.

For executive agencies to preserve their proper role in policy and program implementation, four actions are needed: (1) rebuilding staff capacity, (2) providing dedicated resources for program evaluation and data collection, (3) setting priorities to ensure that information arrives when needed, and (4) ensuring honest, full reporting, both to agency policymakers and to the Congress.

Charles A. Bowsler

Charles A. Bowsler

Contents

Letter	1
Why Program Evaluation Is Important	6
The Current Decline in Evaluative Capability	9
Evaluation Findings in Specific Agencies	12
Rebuilding Capacity	19
If Agencies Don't, Others Will	23
Related GAO Products	26
Transition Series	28

Why Program Evaluation Is Important

“The Department doesn’t have anything on that” is rarely an acceptable answer by federal officials to questions about their programs, yet that is a statement heard with increasing regularity.

We do not minimize the difficulties of determining program effectiveness. We do emphasize that there are proven methods for getting credible answers to questions about program operation and effectiveness. These methods collectively are called program evaluation.

When an Iranian airliner is shot down by an Aegis missile in the Persian Gulf, or when Wall Street goes into economic meltdown, the call for information comes loud and clear. Less dramatic, but equally real, are the “routine” demands for data on how effectively the federal government is using its trillion-dollar-plus budget. In 1988, for example, federal officials testified several thousand times before Congress and provided over 3,000 legislatively required reports, according to agency estimates. At the least, these reports should provide relevant, timely, and technically adequate data on federal programs.

Program evaluation—when it is available and of high quality—provides sound information about what programs are actually

delivering, how they are being managed, and the extent to which they are effective or cost-effective.

Unfortunately, we have discovered through our surveys that program evaluation and the data collection that supports it are—with few exceptions—in a depleted state in executive agencies today. Further, case studies show that basic data are lacking on such disparate and wide-ranging issues as health care quality, the state of the environment, and the results of weapon system testing.

This shortage of evaluative information should be of immediate concern to federal managers: they need it both to run their programs and to justify their decisions and performance. It is certainly of concern to GAO: it means a corresponding overflow in requests to us for information on program results. But the most important point is that the shortage is really everyone's concern, because good information is not just a management tool. It is the responsibility of government to the people of this country, and it is not in a healthy state.

In the past, sound evaluations have contributed strongly to well-informed decisions, such as those to maintain effective programs like Head Start or Runaway and

Homeless Youth. Without such evaluations, these important programs might not have survived the budget cuts that depleted other programs whose effectiveness had not been so solidly established. Also, in the past, programs such as the breeder reactor or binary chemical weapons have been dropped or slowed down because evaluations convincingly showed their extreme weakness.

Today, however, the capability to perform program evaluation is drying up, not the least in such areas as defense, health care, education, and the environment, where it is precisely most needed.

The Current Decline in Evaluative Capability

Program evaluation support for the administration's top managers may be insufficient for their needs. In 1984, and again in 1988, a review of evaluation services in non-Defense agencies found a significant general decline since 1980 in the capacity and availability of data on federal programs, although agencies varied.

Professional staff in agency evaluation units decreased overall by 22 percent, from about 1,500 to about 1,200, between 1980 and 1984, while total staff for the agencies as a whole decreased only 6 percent. A 1988 study of 15 units that had been active in 1980 showed a 52-percent decrease in professional staff, an additional 12 percent since 1984.

Funds for program evaluation decreased by 37 percent between 1980 and 1984 (in constant 1980 dollars), while agency budgets had an overall increase of 4 percent. This reduction in resources slowed in the period 1984-1988, which saw a 6-percent decrease, primarily because of legislative set-asides that accounted for roughly 60 percent of total allocations in both 1984 and 1988. (In contrast, in 1980 the majority of resources came from departmental budgets.) The amount of program evaluation funds needed differs among programs, but overall, 1984 funding was only about

\$111 million to cover all costs of assessing the results of all domestic programs, according to the agencies.

Despite changes in staff and funds, the number of evaluation studies decreased between 1980 and 1984 by only 3 percent. The impression of greater efficiency—same or more work with fewer resources—is misleading. The same work was not being done in 1984 or in 1988 as had been done in 1980. Work shifted from complex evaluations that give more precise measures of program effects to less complex studies and nontechnical reports.

In our 1988 sample of units, there was a greater reliance on external professionals. Unlike earlier years, when internal professionals performed small-scale, quick-turn-around studies, staff shortages appear to have reached a critical level, forcing evaluation units to contract out even small-scale studies. To make matters worse, one agency official suggested that this pool of qualified contractors is also shrinking.

Further, reports are increasingly produced at the request of program managers, and primarily for internal consumption. Evaluations for external consumption—for congressional oversight and public scrutiny—were limited in number and were primarily

studies mandated by Congress, which set
aside funds for this purpose.

Evaluation Findings in Specific Agencies

The concerns just mentioned about the likely quality of individual studies produced within this overall context of evaluation decline were fully borne out by our in-depth examinations of evaluation functions in four agencies. We found problems in the development, dissemination, and use of evaluations that were both consistent and extremely severe.

Inadequacies of DOD Evaluative Data

After unprecedented real growth from fiscal years 1980 to 1985, defense funding has declined and may well decline still further. The new Congress and top management will have to make difficult trade-offs among programs. Unfortunately, they may be lacking the data crucial to these decisions.

The quality of program evaluation information on some defense programs is so low that findings are misleading; in other cases, information is nonexistent; and in still others, data are incomplete. This is especially important because of the multi-billion-dollar price tags on so many DOD programs. Furthermore, adequate systems are not in place for getting sound evaluative information about DOD programs to its own top managers or to Congress. For example, we found

- There were virtually no data on defense industrial production capabilities and problems at the subcontractor level. Collecting our own data, we found many unsuspected problems, including a heavy dependency on foreign suppliers, shortages of production machines and test equipment, and a large number of processes proprietary to individual contractors.
- Program testers failed to collect data on so many crucial questions in the Bigeye chemical bomb's development tests that it is unclear whether the bomb has met its specifications.
- In DOD's Live Fire Testing Program, weapons were often procured based only on computer analyses of vulnerability and lethality estimates, with little or no real data on weapon system effectiveness against actual threats under field conditions.
- The Secretary of Defense and Congress regularly receive information that contains omissions and inaccuracies in the data. They also receive assessments that consistently transmit a more favorable presentation of test adequacy and system performance than is warranted by the facts.

Tough Medicare
Issues

Medicare—second only to Social Security among HHS entitlement programs—expended an estimated \$70 billion, or 10 percent of the non-Defense budget, on health services to the elderly in fiscal year 1987. If current trends continue, rising federal health care costs for the elderly will overtake federal spending for retirement income. Whether efforts to control Medicare costs and reduce unnecessary care are severely damaging access to care and the quality of that care is a major issue for program evaluation today. The need is great, also, to determine whether proposed alternatives are in fact viable solutions.

The Medicare program's evaluation system today cannot support informed decisions. For example

- The Office of Research and Demonstrations—the logical unit within the Health Care Financing Administration to assume responsibility for Medicare program evaluation—has serious evaluation management problems.
- Information is meager on the quality of care that Medicare hospital patients receive.

- Information on quality of care in other settings is virtually nonexistent.
-

Adequacy and
Availability of EPA
Regulatory Data

EPA does not know whether many multibillion-dollar environmental programs are effective in achieving planned goals, because its evaluation information is misleading, inadequate, or incomplete. Shortfalls abound in the basic descriptive information needed to establish and monitor environmental regulations, and the effects of some costly environmental programs that entail considerable compliance burdens are undocumented. For example

- Data on hazardous waste volume and capacity are lacking or suspect.
- EPA has not comprehensively characterized the groundwater contaminants regulated by the states.
- The states lack important data to set technically sound groundwater protection standards.
- Major questions remain unanswered about the quality of the nation's rivers and streams.

- Data do not exist on the effectiveness of EPA's drinking water standards as protection against groundwater contamination, yet many states are using them.
- The effectiveness of the multibillion-dollar wastewater construction grants program in improving water quality has not been adequately evaluated.
- In many program areas, EPA has not conducted program evaluations to determine the health effects of risk-reduction efforts.

Our conclusions in the program evaluation area are consistent with those of the GAO general management review of EPA (RCED-88-101), which emphasized the need for the agency to set measurable objectives and to manage its resources to achieve and assess program and policy goals. On the plus side, we found that some of EPA's scientific risk assessment work has been very well conducted.

Education: C- On
Information

The responsibility for collecting nationally useful information on who is getting educated, how well, and what can be done to improve the excellence of U.S. education is dispersed across evaluation, statistical, and research units throughout the Department of Education. Since the early 1970s, education studies have declined sharply.

Program evaluation studies, in particular, have declined by over 65 percent. Moreover, the statistical data produced were often stale (3-5 years, or more, out of date) and the technical quality varied.

The National Center for Educational Statistics has initiated efforts to upgrade national data bases and has overhauled its management, providing some much-needed stability and technical expertise. These improvements are new, and they will require continued management support over the next few years.

The research function could benefit from the lessons learned in reconstructing the statistical systems. Beginning in 1989, the laboratory and center systems supported by research funds are scheduled for recompetition. Currently they consume about 95 percent of the Department's non-programmatic research resources, yet no evaluative data exist on whether this is an optimal use of limited funds. How priorities for these awards are set will affect the flow of educational information for the next 5 years. For example, will top priority be set on the disadvantaged? on higher education? or on others of the many topics competing for attention?

Much of the Department's program evaluation resources are used for internal analyses of proposed policies. But in part through congressional mandates and set-asides, there are several nationally significant evaluation studies are nearing the end of the pipeline. In 1989, major evaluations of such sensitive topics as bilingual education and educational reform should be completed. In the recent past, there have been criticisms of how the Department reviews and releases this kind of policy-sensitive research. For example, headline-getting press conferences have been held before reports were available for proper professional and public review.

Rebuilding Capacity

Throughout many federal agencies the information pipeline for program oversight and management is drying up. The reduction in staff and funds needed to determine program effectiveness leaves managers unprepared to answer tough questions about program costs and results, and vulnerable to incidents such as the Vincennes situation in the Persian Gulf. Top-management support for collecting this information and insistence on honest reporting are urgently needed.

What Is Needed: Capacity

The infrastructure—the capacity for program evaluation—has generally eroded and will have to be rebuilt. For example

- Staff who are professionally trained in evaluation techniques and who also have the requisite understanding of programs will have to be recruited and trained in federal procedures. Graduate training is usually a minimum requirement for these positions.
- Many evaluation units may be inadequately staffed. As a rough estimate, one evaluator can manage about \$1 million in contract activity and one manager is needed for every six or seven evaluators, depending on the size and complexity of a program.

- Locating at least one evaluation unit at the top-management level is essential, so that priorities can directly reflect agency and congressional concerns and the results can flow without censorship to the top.
- Relationships that support planning and budget functions also need to be strengthened.

What Is Needed:
Resources

A notable imbalance persists in executive branch funding of program evaluation. Some agencies say they do not do any evaluations. In others, such costly programs as Clean Water lack adequate data on program effectiveness. Given the resources that can be saved through sound evaluation and the need for an agency to be able to show that its programs are effective and well managed,

- program evaluation needs to be generally strengthened throughout the government, and
- resources need to be both dedicated to program evaluation and expanded in those agencies where evaluation has been eroded or is not functioning well.

Program evaluation personnel can present a tempting target to budget cutters in times of tight resource constraints, simply

because they are not line staff. This temptation must be resisted.

What Is Needed:
Priorities

Within the four areas we examined in depth, we believe the top-priority evaluation issues for the next few years are:

- integrity of weapon system evaluation,
- long-term medical care needs,
- cost-effectiveness of environmental systems, and
- excellence and competitiveness of U.S. education.

Top leadership in all agencies should review their program evaluation efforts to determine what areas in their programs need highest priority attention.

What Is Needed:
Honest Reporting

Some of the problems with agency evaluations have included the failure to conduct necessary studies, nontechnical influence on draft reports that have concealed or distorted findings, technical flaws affecting study quality, the uncertain access of top managers to complete and unvarnished study findings, and the limited use of evaluations in making policy.

Needed improvements in the reporting and use of program evaluation include

- assurance that objective data get to top decisionmakers. The review process that helps ensure technical adequacy and balance should not deter timeliness or candor.
- more extensive communication between requesters and evaluators in the early, middle, and late stages of a study. Such communication, which includes evaluator participation in the early formulation of data requests, is lacking, particularly in studies mandated by Congress. Communication is crucial to keeping results of costly studies off some dusty shelf and putting them onto agendas for action.
- greater attention to the prospective aspect of evaluations, before new programs, policies, or regulations are launched. "Front-end" evaluations can prevent a poor use of funds and target resources where they are most likely to be effective.

If Agencies Don't, Others Will

A consequence of the drawdown in evaluation capacity is that agencies may find themselves fenced off from evaluation-based debates about their own programs. That is, not only may agency-generated information for the public be lacking, but agencies themselves may not have the data needed to be most convincing about what they think should be started, what they think should be stopped, and what they think should be changed. The terms of the debates and the data brought to bear on these issues may be someone else's call.

Further, the problems we have described mean that to a growing extent, even GAO and others that often draw on agency records and data may be able to report little more than "Information to answer this key question is not available." That is, the loss of program evaluation information has repercussions that go beyond whether an agency head has to testify, "My Department does not have any data on that point," to whether a wide range of groups that rely in part on federally collected evaluations can independently tell Congress what is happening in the executive branch.

Over the years, the congressional appetite for the data needed for oversight on complex questions about the operations and

consequences of federal programs has not slackened. Congress increasingly is relying on GAO and its sister agencies—the Congressional Budget Office (CBO), the Office of Technology Assessment (OTA), and the Congressional Research Service (CRS)—to do studies that might appropriately be conducted by executive branch agencies. We have, for example, been tasked legislatively with a major set of analyses on the effect of immigration reform and with determination of the numbers of homeless children and youth.

Clearly, if the executive branch cannot provide timely, relevant, technically adequate, and credible information on the programs that it is responsible for administering, Congress will continue to write us into legislation that mandates these important studies.

Our mission, of course, is to provide credible information to Congress and help ensure that Congress is not limited to reports from special interest, public interest, or other groups. However, although Congress may have the high ground on information, we should not, and indeed cannot, do it all. Such a role exceeds our resources. Moreover, it could lead to a serious imbalance between the branches on who calls the shots on the information that

has become a vital influence in major debates on national policy and that is considered an important indicator of a credible government.

Each agency head should have, as a top priority, an honest inventory of what information is coming on line, whether it will be available in time to affect key decisions over the next 4 years, what the technical quality and relevance of the evaluations are, and what gaps need to be plugged first. In a few instances, the situation may be relatively good; but in many others, we must emphasize that a renewed commitment to program evaluation is urgently needed.

Related GAO Products

Weapons Testing: Quality of DOD Operational Testing and Reporting (PEMD-88-32BR, July 26, 1988).

HCFA Research: Agency Practices and Other Factors Threaten Quality of Mandated Studies (PEMD-88-9, June 3, 1988).

DOD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results (PEMD-88-3, Dec. 29, 1987).

Education Information: Changes in Funds and Priorities Have Affected Production and Quality (PEMD-88-4, Nov. 4, 1987).

Live Fire Testing: Evaluating DOD's Programs (PEMD-87-17, Aug. 17, 1987).

Hazardous Waste: Uncertainties of Existing Data (PEMD-87-11BR, Feb. 18, 1987).

Federal Evaluation: Fewer Units, Reduced Resources, Different Studies From 1980 (PEMD-87-9, Jan. 23, 1987).

Water Quality: An Evaluation Method for the Construction Grants Program—A Methodology, Vol. 1; Case Studies, Vol. 2 (PEMD-87-4A and 4B, Dec. 17, 1986).

The Nation's Water: Key Unanswered Questions About the Quality of Rivers and Streams (PEMD-86-6, Sept. 19, 1986).

Post-Hospital Care: Efforts to Evaluate Medicare Prospective Payment Effects Are Insufficient (PEMD-86-10, June 2, 1986).

Bigeye Bomb: An Evaluation of DOD's Chemical and Development Tests (PEMD-86-12BR, May 23, 1986).

Assessing Production Capabilities and Constraints in the Defense Industrial Base (PEMD-85-3, Apr. 4, 1985).

Transition Series

The Budget Deficit (GAO/OCG-89-1TR)

The Public Service (GAO/OCG-89-2TR)

Revenue Options (GAO/OCG-89-3TR)

Financial Services Industry Issues (GAO/OCG-89-4TR)

International Trade Issues (GAO/OCG-89-5TR)

Information Technology Issues (GAO/OCG-89-6TR)

Financial Management Issues (GAO/OCG-89-7TR)

Program Evaluation Issues (GAO/OCG-89-8TR)

Defense Issues (GAO/OCG-89-9TR)

Health and Human Services Issues (GAO/OCG-89-10TR)

Commerce Issues (GAO/OCG-89-11TR)

Agriculture Issues (GAO/OCG-89-12TR)

Justice Issues (GAO/OCG-89-13TR)

Veterans Affairs Issues (GAO/OCG-89-14TR)

NASA Issues (GAO/OCG-89-15TR)

Energy Issues (GAO/OCG-89-16TR)

Treasury Issues (GAO/OCG-89-17TR)

Education Issues (GAO/OCG-89-18TR)

Department of State Issues (GAO/OCG-89-19TR)

Environmental Protection Agency Issues (GAO/
OCG-89-20TR)

Department of Labor Issues (GAO/OCG-89-21TR)

Transition Series

Housing and Urban Development Issues (GAO/
OCG-89-22TR)

Foreign Economic Assistance Issues (GAO/
OCG-89-23TR)

Interior Issues (GAO/OCG-89-24TR)

Transportation Issues (GAO/OCG-89-25TR)

Internal Revenue Service Issues (GAO/OCG-89-26TR)