

May 2003

PROGRAM
EVALUATION

An Evaluation Culture
and Collaborative
Partnerships Help
Build Agency Capacity



G A O

Accountability * Integrity * Reliability

Why GAO Did This Study

Agencies are increasingly asked to demonstrate results, but many programs lack credible performance information and the capacity to rigorously evaluate program results. To assist agency efforts to provide credible information, GAO examined the experiences of five agencies that demonstrated evaluation capacity in their performance reports: the Administration for Children and Families (ACF), the Coast Guard, the Department of Housing and Urban Development (HUD), the National Highway Traffic Safety Administration (NHTSA), and the National Science Foundation (NSF).

PROGRAM EVALUATION

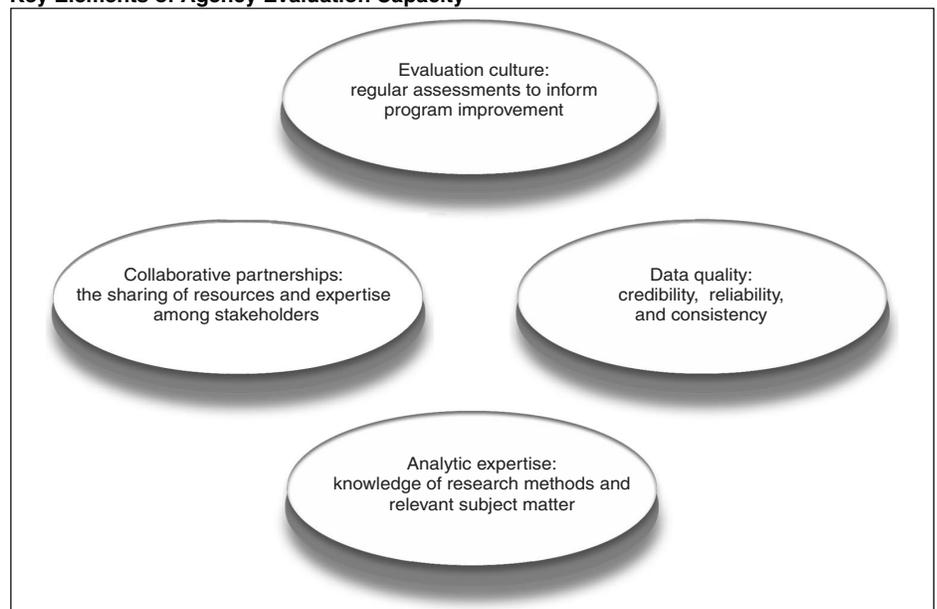
An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity

What GAO Found

In the five agencies GAO reviewed, the key elements of evaluation capacity were an evaluation culture—a commitment to self-examination, data quality, analytic expertise, and collaborative partnerships. ACF, NHTSA, and NSF initiated evaluations regularly, through a formal process, while HUD and the Coast Guard conducted them as specific questions arose. Access to credible, reliable, and consistent data was critical to ensure findings were trustworthy. These agencies needed access to expertise in both research methods and subject matter to produce rigorous and objective assessments. Collaborative partnerships leveraged resources and expertise. ACF, HUD, and NHTSA primarily partnered with state and local agencies; the Coast Guard partnered primarily with federal agencies and the private sector.

The five agencies used various strategies to develop and improve evaluation: Commitment to learning from evaluation developed to support policy debates and demands for accountability. Some agencies improved administrative systems to improve data quality. Others turned to specialized data collection. All five agencies typically contracted with experts for specialized analyses. Some agencies provided their state partners with technical assistance. These five agencies used creative strategies to leverage resources and obtain useful evaluations. Other agencies could adopt these strategies—with leadership commitment—to develop evaluation capacity, despite possible impediments: constraints on spending, local control over flexible programs, and restrictions on federal information collection. The agencies agreed with our descriptions of their programs and evaluations.

Key Elements of Agency Evaluation Capacity



Source: GAO.

www.gao.gov/cgi-bin/getrpt?GAO-03-454.

To view the full report, including the scope and methodology, click on the link above. For more information, contact Nancy Kingsbury at (202) 512-2700 or KingsburyN@gao.gov.

Contents

Letter	1
Results in Brief	2
Background	3
Scope and Methodology	5
Case Descriptions	6
Key Elements of Evaluation Capacity	9
Strategies for Enhancing Evaluation Capacity	14
Factors That Impede Building Evaluation Capacity	24
Observations	24
Agency Comments	25
Bibliography	26
Related GAO Products	28
Figures	
Figure 1: Key Elements of Agency Evaluation Capacity	9
Figure 2: Agency Strategies for Building Evaluation Capacity	15

Abbreviations

ACF	Administration for Children and Families
AFDC	Aid to Families with Dependent Children
ASPE	Assistant Secretary for Planning and Evaluation
CDBG	Community Development Block Grant
COV	Committee of Visitors
CPD	Community Planning and Development
DOT	Department of Transportation
FARS	Fatality Analysis Reporting System
GPRA	Government Performance and Results Act of 1993
HHS	Department of Health and Human Services
HOME	HOME Investment Partnerships Program
HUD	Department of Housing and Urban Development
JOBS	Job Opportunities and Basic Skills Training
MDRC	Manpower Demonstration Research Corporation
MIS	management information system
MPA	Masters in Public Administration
NHTSA	National Highway Traffic Safety Administration
NSF	National Science Foundation
OMB	Office of Management and Budget
ONDCP	Office of National Drug Control Policy
PART	Program Assessment Rating Tool
PD&R	Office of Policy Development and Research
TANF	Temporary Assistance for Needy Families

This is a work of the U.S. Government and is not subject to copyright protection in the United States. It may be reproduced and distributed in its entirety without further permission from GAO. It may contain copyrighted graphics, images or other materials. Permission from the copyright holder may be necessary should you wish to reproduce copyrighted materials separately from GAO's product.



G A O

Accountability * Integrity * Reliability

United States General Accounting Office
Washington, DC 20548

May 2, 2003

The Honorable Susan Collins
Chairman
Committee on Governmental Affairs
United States Senate

The Honorable George Voinovich
Chairman
The Honorable Richard Durbin
Ranking Minority Member
Subcommittee on Oversight of Government Management,
the Federal Workforce, and the District of Columbia
Committee on Governmental Affairs
United States Senate

The Honorable Tom Davis
Chairman
Committee on Government Reform
House of Representatives

Federal agencies are increasingly expected to focus on achieving results and to demonstrate, in annual performance reports and budget requests, how their activities help achieve agency or governmentwide goals. The current administration has made linking budgetary resources to results one of the top five priorities of the President's Management Agenda. As part of this initiative, the Office of Management and Budget (OMB) has begun to rate agency effectiveness through summarizing available performance and evaluation information. However, in preparing the 2004 budget, OMB found that half the programs they rated were unable to demonstrate results. We have also noted limitations in the quality of agency performance and evaluation information and agency capacity to produce rigorous evaluations of program effectiveness.¹ To sustain a credible performance-based focus in budgeting and ensure fair assessments of agency and program effectiveness, federal agencies, as

¹U.S. General Accounting Office, *Performance Budgeting: Opportunities and Challenges*, GAO-02-1106T (Washington, D.C.: Sept. 19, 2002).

well as those third parties that implement federal programs, will require significant improvements in evaluation information and capacity.

To assist agency efforts to provide credible information on program effectiveness, we (1) reviewed the experiences of five agencies with diverse purposes that have demonstrated evaluation capacity—the ability to systematically collect, analyze, and use data on program results and (2) identified useful capacity-building strategies that other agencies might adopt. The five agencies are the Administration for Children and Families (ACF), the Coast Guard, the Department of Housing and Urban Development (HUD), the National Highway Traffic Safety Administration (NHTSA), and the National Science Foundation (NSF). We developed this report under our own initiative, and are addressing this report to you because of your interest in encouraging results-based management.

To identify the five cases, we reviewed agency documents and evaluation studies for examples of agencies incorporating the results of program evaluations in annual performance reports. We selected these five cases because they include diverse program purposes: regulation, research, demonstration, and service delivery (directly or through third parties). We reviewed agency evaluation studies and other documents and interviewed agency officials to identify (1) the key elements of each agency’s evaluation capacity and how they varied across the agencies and (2) the strategies these agencies used to build evaluation capacity.

Results in Brief

In the agencies we reviewed, the key elements of evaluation capacity were: an evaluation culture, data quality, analytic expertise, and collaborative partnerships. Agencies demonstrated an evaluation culture through regularly evaluating how well programs were working. Managers valued and used this information to test out new initiatives or assess progress toward agency goals. Agencies emphasized access to data that were credible, reliable, and consistent across jurisdictions to ensure that evaluation findings were trustworthy. Agencies also needed access to analytic expertise to produce rigorous and objective assessments at either the federal or another level of government. Each agency needed research expertise, as well as expertise in the relevant program field, such as labor economics, or engineering. Finally, agencies formed collaborations with program partners and others to leverage resources and expertise to obtain performance information.

The key elements of evaluation capacity took various forms and were more or less apparent across the five cases we reviewed. At ACF, NHTSA,

and NSF, the evaluation culture was readily visible because these agencies initiated evaluations on a regular basis, through a formal process. In contrast, at HUD and the Coast Guard, evaluations were conducted on an ad hoc basis, in response to questions raised about specific initiatives or issues. At ACF, HUD, and NHTSA, where states and other parties had substantial control over the design and implementation of the program, access to credible data played a critical role, and partnerships with state and local agencies were more evident. At the Coast Guard, partnerships with federal agencies and the private sector were more evident.

The five agencies we reviewed used various strategies to develop and improve evaluation. Agency evaluation culture, an institutional commitment to learning from evaluation, was developed to support policy debates and demands for accountability. Some agencies developed their administrative systems to improve data quality for evaluation. Others turned to special data collections. To ensure common meaning of data collected across localities, some agencies created specialized data systems. The five federal agencies typically contracted with experts for specialized analyses. These agencies also helped states obtain expertise through developing program staff or hiring local contractors. Some collaborative partnerships developed naturally through pursuit of common goals, while other agencies actively solicited their stakeholders' involvement in evaluation.

To provide credible information on program effectiveness, these five agencies described creative strategies for leveraging their resources and those of their program partners. Supported by leadership commitment, other agencies could adopt these strategies to develop evaluation capacity. However, agency officials also cited conditions that can be expected to create impediments for others as well: constraints on spending program resources on oversight, local control over the design and implementation of flexible programs, and restrictions on federal information collection.

Background

Federal agencies are increasingly expected to demonstrate effectiveness in achieving agency or governmentwide goals. The Government Performance and Results Act of 1993 (GPRA) requires federal agencies to report annually on their progress in achieving agency and program goals. The President's Budget and Performance Integration initiative extends GPRA's efforts to improve government performance and accountability by

bringing performance information more directly into the budgeting process.² In developing the fiscal year 2004 budget, OMB (1) asked agencies to more directly link expected performance with requested program activity funding levels and (2) prepared effectiveness ratings, with a newly devised Program Assessment Rating Tool (PART), for about one-fifth of federal programs.

The PART consists of a standard set of questions that OMB and agency staff complete together, drawing on available performance and evaluation information. The PART questions assess the clarity of program design and strategic planning and rate agency management and program performance. The PART asks, for example, whether program long-term goals are specific, ambitious, and focused on outcomes, and whether annual goals demonstrate progress toward achieving long-term goals. It also asks whether the program has achieved its annual performance goals and demonstrated progress toward its long-term goals. Ratings are designed to be evidence-based, drawing on a wide array of information, including authorizing legislation, GPRA strategic plans and performance plans and reports, financial statements, Inspector General and our reports, and independent program evaluations.

Almost a decade after GPRA was enacted, the accuracy and quality of evaluation information necessary to make the judgments called for in rating programs is highly uneven across the federal government. GPRA expanded the supply of results-oriented performance information generated by federal agencies. However, in the 2004 budget, OMB rated 50 percent of the programs evaluated as “Results Not Demonstrated” because they did not have adequate performance goals or had not collected data to produce evidence of results. We have noted that agencies have had difficulty assessing (1) many program outcomes that are not quickly achieved or readily observed and (2) contributions to outcomes that are only partly influenced by federal funds.³ To help explain the linkages between program activities, outputs and outcomes, a program evaluation—depending on its focus—may review aspects of program operations or factors in the program environment. In impact evaluation, scientific research methods are used to establish a causal connection

²Strategic management of human capital, competitive sourcing, improving financial performance, and expanded electronic government are the other four initiatives in the President’s Management Agenda, described at the Web site www.results.gov.

³[GAO-02-1106T](#).

between program activities and outcomes and to isolate the program's contributions to them. Our previous work raised concerns about the capacity of federal agencies to produce evaluations of program effectiveness.⁴ Few deployed the rigorous research methods required to attribute changes in underlying outcomes to program activities. Yet, we have also seen how some agencies have profitably drawn on systematic program evaluations to explain the reasons for program performance and identify strategies for improvement.⁵

Scope and Methodology

To identify ways that agencies can improve evaluation capacity, we conducted case studies of how five agencies had built evaluation capacity over time. To select the cases, we reviewed departmental and agency performance plans and reports, as well as evaluation reports, for examples of how agency performance reports had incorporated evaluation results. To obtain a broadly applicable set of strategies, we selected cases to reflect a diversity of federal program purposes. Because program purpose is central to considering how to evaluate effectiveness or worth, the type of evaluation an agency conducts might shape the key elements of the agency's evaluation capacity. For this review, we selected cases based on a classification of program purposes employed in our previous study—demonstration, regulation, research, and service delivery.⁶

The first three classifications are represented in our case selection of ACF, NHTSA, and NSF. For service delivery, we chose one agency that delivers services directly to the public (the Coast Guard), and another that provides services through third parties (HUD). Although we selected cases to capture a diversity of federal program experiences, the cases should not be considered to represent all the challenges faced or strategies used. We describe all five cases in the next section.

⁴U.S. General Accounting Office, *Program Evaluation: Agencies Challenged by New Demand for Information on Program Results*, [GAO/GGD-98-53](#) (Washington, D.C.: Apr. 24, 1998).

⁵U.S. General Accounting Office, *Program Evaluation: Studies Helped Agencies Measure or Explain Program Performance*, [GAO/GGD-00-204](#) (Washington, D.C.: Sept. 29, 2000).

⁶U.S. General Accounting Office, *Program Evaluation: Improving the Flow of Information to the Congress*, [GAO/PEMD-95-1](#) (Washington, D.C.: Jan. 30, 1995). Demonstration programs are defined here as those that aim to produce evidence of the feasibility or effectiveness of a new approach or practice. Other program types include statistical, acquisition, and credit programs.

For each agency, to identify the key elements of evaluation capacity and strategies used to build capacity, we reviewed agency and program materials and interviewed agency officials. Our findings are limited to the examples reviewed and do not necessarily reflect the full scope of each agency's evaluation activities. For example, we did not review all HUD evaluations, only evaluations of flexible grant programs. We conducted our work between June 2002 and March 2003 in accordance with generally accepted government auditing standards.

We requested comments on a draft of this report from the heads of the agencies responsible for the five cases. The Departments of Health and Human Services and Housing and Urban Development provided technical comments that we incorporated where appropriate throughout the report.

Case Descriptions

We describe the program structures, major activities, and evaluation approaches for the five cases in this section.

Administration for Children and Families (ACF)

ACF, in the Department of Health and Human Services (HHS), oversees and helps finance programs to promote the economic and social well-being of families, individuals, and communities. Through the Temporary Assistance for Needy Families (TANF) program, ACF provides block grants to states so that they can develop programs of financial and other assistance. These programs help needy families find employment and economic self-sufficiency. In 1996, TANF replaced Aid to Families with Dependent Children (AFDC), commonly referred to as welfare, and the Job Opportunities and Basic Skills Training (JOBS) programs. Under the AFDC program, states conducted demonstrations, for three decades, to test out alternative approaches for moving recipients off welfare and into work. As part of a broad array of studies of poverty populations and programs, ACF and the Office of the Assistant Secretary for Planning and Evaluation (ASPE) continue to support evaluations of state welfare-to-work experiments, including implementation and process studies, as well as impact studies based on experimental evaluation methods.

Coast Guard

In the Department of Transportation (DOT), the Coast Guard provides diverse customer services to ensure safe and efficient marine transportation, protect national borders, enforce maritime laws and treaties, and protect natural resources. The Coast Guard's mission includes enhancing mobility, by providing aids to navigation, icebreaking services, bridge administration, and vessel traffic management activities; security, through law enforcement and border control activities; and

safety, through programs for accident prevention, response, and investigation. The agency monitors numerous indicators to assess allocation of resources to and performance in achieving service goals. The Coast Guard has initiated an effort to evaluate its direct services and resource-building efforts through a Readiness Management System, which covers people, equipment, and stations. In addition, special studies of the success of specific initiatives may be contracted out.

Housing and Urban Development (HUD)

The HUD Office of Community Planning and Development (CPD) provides financial and technical assistance to states and localities in order to promote community-based efforts to develop housing and economic opportunities. CPD's largest program, the Community Development Block Grant program (CDBG) has, for the past two decades, provided formula grants to cities, urban counties, and states to foster decent, affordable housing, and expanded economic opportunities for low- and moderate-income people. Communities may use funds for a wide range of activities directed toward neighborhood revitalization, economic development, and improved community facilities and services.⁷ CPD also administers the HOME Investment Partnerships Program (HOME), a block grant to state and local governments, to create decent, affordable housing for low-income families. First funded in 1992, HOME has more specific goals than CDBG: (1) to help build, buy, or rehabilitate affordable housing for rent or home ownership or (2) to provide direct tenant-based rental assistance. In addition to maintaining information on housing need, market conditions, and programs across the department, HUD's Office of Policy Development and Research (PD&R) supports studies of the use and benefits of the CDBG and HOME grants.

National Highway Traffic Safety Administration (NHTSA)

To promote highway safety, DOT's NHTSA develops regulations and provides financial and technical assistance to states and local communities. These communities, in turn, conduct highway safety programs that respond to local needs. To identify the most effective and efficient means to bring about safety improvements, NHTSA also conducts research and development in vehicle design and driver behavior. To assess the effectiveness of its regulatory and safety promotion efforts, NHTSA

⁷CDBG programs are often small-scale "bricks and mortar" initiatives that may include such activities, among others, as the reconstruction of streets, water and sewer facilities, and neighborhood centers, and rehabilitation of public and private buildings.

reviews outcomes, such as reduction of alcohol-related fatalities or increase in helmet or safety belt use. To illuminate the causes and outcomes of crashes and evaluate safety standards and initiatives, NHTSA analyzes state and specially created national databases, for example, the Fatality Analysis Reporting System (FARS).

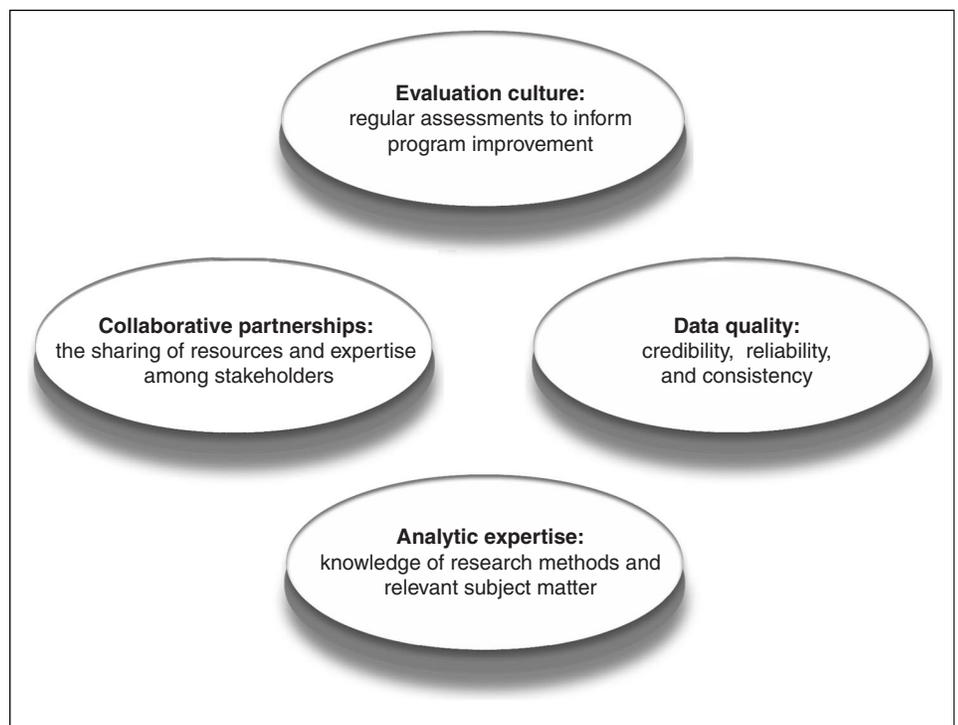
**National Science
Foundation (NSF)**

NSF funds education programs and a broad array of research projects in the physical, geological, biological, and social sciences; mathematics; computing; and engineering; which are expected to lead to innovative discoveries. NSF provides support for investigator-initiated research proposals that are competitively selected, based on merit reviews. The agency has a long-standing review infrastructure in place: for each individual research program, panels of outside experts rank proposals on merit. NSF also convenes panels of independent experts as external advisers—a Committee of Visitors (COV)—to peer review the technical and managerial stewardship of a specific program or cluster of programs periodically, compare plans with progress made, and evaluate outcomes to determine whether the research contributes to NSF mission and goals. Each COV, based on an academic peer review model, usually consists of 5 to 20 external experts, who represent academia, industry, government, and the public sector. These reviews serve as a means of quality assurance for NSF management. About a third of the 220 NSF programs are evaluated each year so that a complete assessment of programs can be accomplished over a 3-year period.

Key Elements of Evaluation Capacity

Four main elements of evaluation capacity were apparent across the diverse array of agencies we reviewed, although they took varied forms. These elements include an evaluation culture, data quality, analytic expertise, and collaborative partnerships. (See figure 1.) Agencies demonstrated an evaluation culture through commitment to self-examination and learning through experimentation. Data quality and analytic expertise were key to ensuring the credibility of evaluation results and conclusions. Agency collaboration with federal and other program partners helped leverage resources and expertise for evaluation.

Figure 1: Key Elements of Agency Evaluation Capacity



Source: GAO.

An Evaluation Culture

Three of our cases—ACF, NHTSA, and NSF—clearly evidenced an evaluation culture: they had a formal, regular process in place to plan, execute, and use information from evaluations. They described a commitment to learning through analysis and experimentation. HUD and

the Coast Guard had more ad hoc arrangements in place when questions about specific initiatives or issues created the demand for evaluations. HUD officials described an annual, consultative process to decide which studies to undertake within budgeted resources.

At ACF, evaluations of state welfare-to-work demonstration programs are a part of a network of long-term federal, state, and local efforts to develop effective welfare policy. Over the past three decades, ACF has supported evaluations of state experiments in how to help welfare recipients find work and achieve economic self-sufficiency. Until TANF replaced AFDC in 1996, states were permitted waivers of federal rules to test new welfare-to-work initiatives on condition that states rigorously evaluate the effects of those demonstrations. Lessons from these evaluations informed not only state policies, but also the formulation of the JOBS work support program in 1988 and the TANF work requirements in 1996. ACF and ASPE continue to support rigorous evaluation of state policy experiments to obtain credible evidence on their effectiveness.

At NHTSA, evaluation was a natural part of meeting the agency's principal responsibility to develop and oversee federal regulations to enhance safety. NHTSA officials said regulatory programs are inherently evaluative in nature because only thorough evaluations of safety issues can lay the foundation for effective regulatory policies. Officials described a tri-part process for evaluation: First, studies to identify the nature of the problem and possible solutions precede proposals for regulatory or other policy changes. Second, cost-benefit analyses identify the expected consequences of alternative approaches. Third, follow-up studies to assess the consequences of regulatory changes are important because effects of some safety innovations may not manifest until 5 or more years after the introduction of changes. These evaluations address the long-term practical consequences of new regulations. At NHTSA, diverse evaluation studies played an integral role throughout the regulatory process.

At NSF, efforts to evaluate its research programs are described as congruent with the scientific community's natural tendency toward self-examination. The NSF oversight body, the National Science Board, issued a report noting that today's environment requires effective management of the federal portfolio of long-term investments in research, including a sustained advisory process that incorporates participation by the science and engineering communities. The COV process to oversee NSF research portfolios has been in place for the past 25 years. During that time, NSF has repeatedly assessed and improved the COV process. COV review templates include questions that assess how the research is contributing to

NSF process and outcome goals. The templates assess, for example, (1) both the integrity and efficiency of the proposal review process and (2) whether the portfolio of projects has made significant contributions to NSF's strategic outcome goals such as "enabling discoveries that advance the frontiers of science, engineering, and technology." Division directors consider COV recommendations in guiding program direction and report on implementation when the COV returns 3 years later.

Data Quality

Credible information is essential to drawing conclusions about program effectiveness. In the cases we examined, agencies strived to ensure the trustworthiness of data obtained through monitoring or evaluation. Data quality involves data credibility and reliability, as well as consistency across jurisdictions. Reliance on states and localities for data on program performance made this a major issue at ACF, HUD, and NHTSA.

For example, NHTSA has devoted considerable effort to develop a series of comparable statistics, on various crash outcomes and safety measures of continuing interest, from varied public and private sources. NHTSA currently maintains seven different public use data files that are updated on a regular (typically, annual) basis.⁸ These data files provide the empirical basis for evaluating NHTSA regulatory programs focused on public health and safety. Although the databases have acknowledged shortcomings, a NHTSA official noted, "These are the most used databases in the world." They are well accepted and used in many program evaluations by safety experts and industry analysts, he noted. NHTSA's record of building well-accepted databases on crash outcomes provides an example of how quality outcome measures can be obtained when causal relationships are well-studied and relatively straightforward.

Analytic Expertise

The agencies reviewed sought access to analytic expertise to ensure assessments of program results would be systematic, credible, and objective. To obtain rigorous analyses, agencies engaged people with research expertise and subject matter expertise to ensure the appropriate interpretation of study findings.

⁸These seven data files provide the empirical basis for analyses of patterns and trends in (1) motor vehicle fatalities; (2) vehicular crashworthiness; (3) medical and financial outcomes of highway crashes; (4) consumer complaints related to vehicles, tires, and other equipment; (5) outcomes of safety defect investigations; (6) motor vehicle compliance testing results; and (7) motor vehicle safety defect recalls.

At ACF, officials indicated that experience in conducting field experiments was critical to obtaining rigorous evaluations. Rigorous methods are required to estimate the net impact of welfare-to-work programs because many other factors, such as the economy, can influence whether welfare recipients find employment. Without similar information on a control group not subject to the intervention, it is difficult to know how many program participants might otherwise have found employment without the program. Conducting a rigorous impact evaluation—randomly assigning cases to either an experimental or control group, tracking the experiences of both groups, and ensuring standardized data collection and appropriate analysis procedures—requires special expertise in social science research. According to ACF officials, they had success in obtaining many such evaluations, in part, because of the existence of a large community of knowledgeable and experienced researchers in universities and contracting firms.

NSF relied on external expert review in its evaluation of research proposals, as well as completed research and development projects. The expert or peer review model allows NSF to tap the specialized knowledge—across many fields—that is critical to assessing whether funded research is making a contribution to the field. Although all agencies required research expertise as well as subject matter expertise that pertained to the program, NSF’s task was compounded by having to cover a broad array of scientific disciplines. Because of the potential for subjectivity in these qualitative judgments, an additional independent review may be necessary to determine the validity of assessments made about progress in achieving scientific discoveries. NSF contracted with PricewaterhouseCoopers, LLP, a professional services organization that provides assurance on the financial performance and operations of business, to independently assess NSF performance results by examining COV scores and justifications.

Collaborative Partnerships

Agencies engaged in collaborative partnerships for the purpose of leveraging resources and expertise. These partnerships played an important role in obtaining performance information. Many agencies share goals with others. Moreover, evaluation capacity at the federal level often depends on the willingness of state and local agencies to participate in rigorous evaluation because of their responsibility for designing and implementing programs. At ACF and HUD, collaboration with both states and localities, as well as with the policy analysis and research communities, plays a central role in evaluation.

Particularly for the Coast Guard, the challenge of achieving national preparedness requires the federal government to form collaborative partnerships with many entities. The primary means of coordination at many ports are port security committees, which offer a forum for federal, state, and local government, as well as private stakeholders to share information and work together collaboratively to make decisions. The breadth of the Coast Guard's public safety responsibilities seemed to increase the number and importance of its partnerships. In order to improve maritime security worldwide, the Coast Guard is working with the International Maritime Organization. Such partnerships can be critical to gaining the resources, expertise, and cooperation of those who must implement the security measures.

In addition, agencies recognized that by working together they could more comprehensively address evaluations of programs. For example, for drug interdiction, the Coast Guard is a key player in deterring the flow of illegal drugs into the United States. For maritime drug interdiction, it is the lead federal agency; it shares responsibility for air interdiction with the U.S. Customs Service. To reduce the illegal drug supply, the Coast Guard coordinates closely with other federal agencies and countries within a Transit Zone⁹ so as to disrupt and deter the flow of illegal drugs. Recognizing the interdependence of agency efforts, the Coast Guard and U.S. Customs Service, along with the Office of National Drug Control Policy (ONDCP), jointly funded a study to examine the deterrence effect of drug enforcement operations on drug smuggling. The study assessed whether interdiction operations or events affected cocaine trafficking.

At ACF and HUD, collaboration with state and local agency program partners was important in evaluating programs. Because of the flexibility in program design given to the states, the studies of flexible grant programs tend to evaluate the effectiveness of a particular state or locality's program, rather than the national program. As an evaluation partner, state agencies need to be willing to participate in rigorous evaluation design and take the risk that programs may not be found to be as successful as they had hoped. While researchers may be hired to design and execute the evaluation, the state agency may be expected to design an innovative program, ensure the program is carried out as planned,

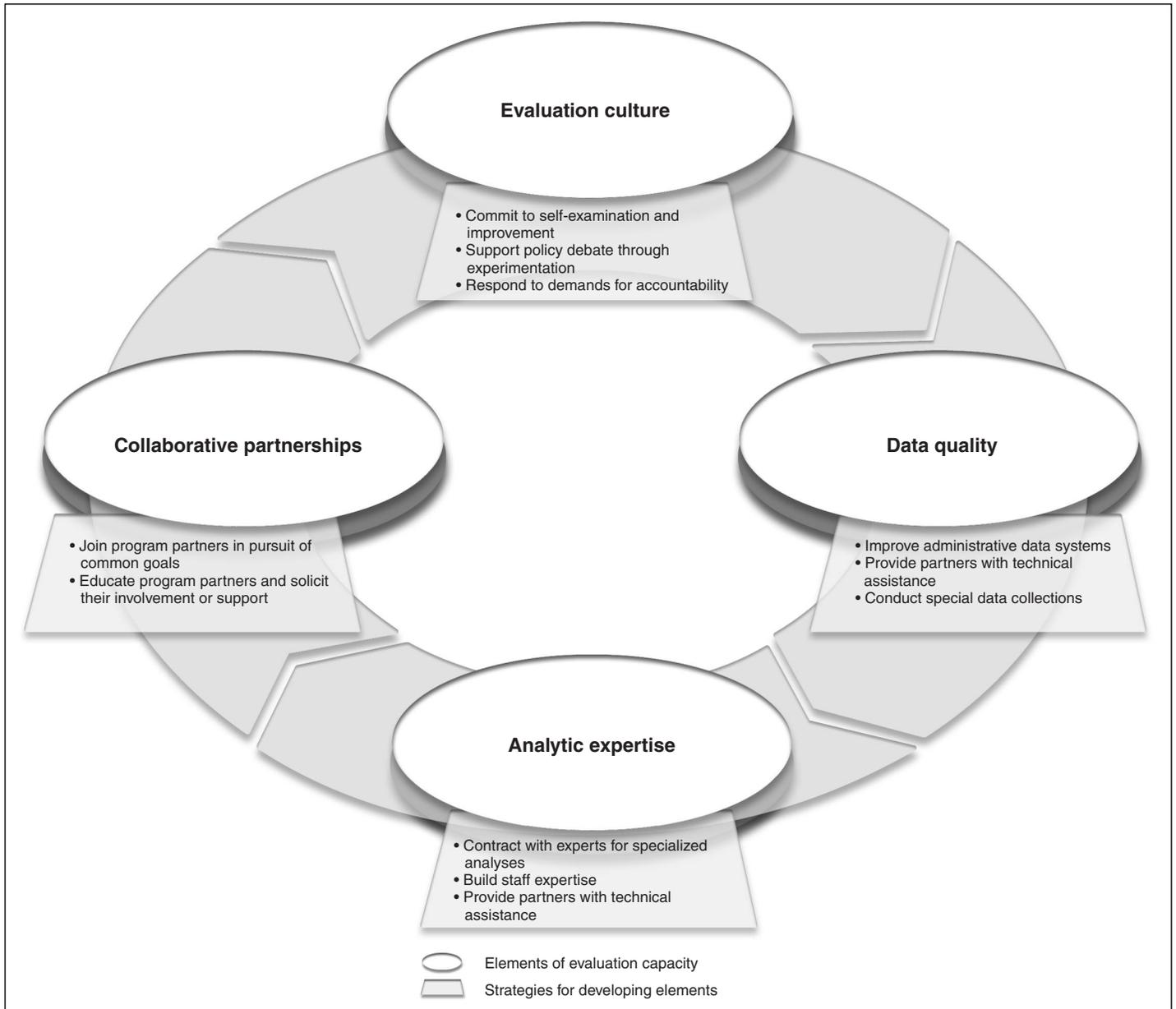
⁹The Transit Zone is a 6 million square mile area, including the Caribbean, Gulf of Mexico, and Eastern Pacific Ocean.

maintain distinctions between the treatment and comparison groups, and ensure collection of valid and reliable data.

Strategies for Enhancing Evaluation Capacity

Through a number of strategies, the five agencies we reviewed developed and maintained a capacity to produce and use evaluations. First, agency managers sustained a commitment to accountability and to improving program performance—to institutionalize an evaluation culture. Second, they improved administrative systems or turned to special data collections to obtain better quality data. Third, they sought out—through external sources or development of staff—whatever expertise was needed to ensure the credibility of analyses and conclusions. Finally, to leverage their evaluation resources and expertise, agencies engaged in collaborations or actively educated and solicited the support and involvement of their program partners and stakeholders. (See figure 2.)

Figure 2: Agency Strategies for Building Evaluation Capacity



Source: GAO.

Institutionalizing an Evaluation Culture

Demand for information on what works stimulated some agencies to develop an institutional commitment to evaluation. The agencies we reviewed did not appear to deliberately set out to build an evaluation culture. Rather, a systematic, reinforcing process of self-examination and improvement seemed to grow with the support and involvement of agency leadership and oversight bodies. ACF and Coast Guard officials described the process as a response to external conditions—policy debates and budget constraints, respectively—that stimulated a search for a more effective approach than in the past.

The evaluation culture at ACF grew as a result of a reinforcing cycle of rigorous research providing credible, relevant information to policymakers who then came to support and encourage additional rigorous research. In the late 1960s, federal policymakers turned to applied social research experiments (for example, the New Jersey-Pennsylvania Negative Income Tax experiment) to inform the debate about how to shape an effective antipoverty strategy. In 1974, the Ford Foundation joined with several federal agencies to set up a nonprofit firm (the Manpower Demonstration Research Corporation (MDRC)) to develop and evaluate promising demonstrations of interventions to assist low-income populations. MDRC's subsequent National Supported Work Demonstration included a rigorous experimental research design that found the interventions did not work; nonexperimental evaluations of similar state programs yielded inconclusive results. A provision permitting waiver of federal rules on condition that states rigorously evaluate those demonstrations—referred to as section 1115 waivers—laid the framework for the next generation of welfare experiments. Results of these demonstrations helped shape the provisions of the JOBS program, enacted in 1988, and a new generation of state experiments that, in turn, shaped the 1996 reforms.

In contrast, Coast Guard officials described their relatively recent development of evaluation capacity as an outgrowth of operational self-examinations, conducted in response to budget constraints. They explained that steep budget cuts in the mid-1990s led the Coast Guard to adopt self-assessments for feedback information on how effectively the agency was using resources, under Total Quality Management initiatives. More recently, the impetus for program evaluation stemmed from the emphasis placed on assessing and improving results in GPRA and the President's Management Agenda. According to Coast Guard officials, they now view the evaluation of program and unit performance as “good business.” Having systems in place that can furnish the necessary trend data has been particularly useful, they said, in supporting and negotiating budget requests. These systems allow the agency to forecast what level of

performance, under different budget scenarios, appropriations committees might expect. The trend data also allow for assessing performance goals and planning program evaluations where performance improvement is needed.

NSF applied the same basic approach it takes to assessing the promise of research proposals to evaluating the quality of completed research programs. NSF described revising the COV process over time, fine-tuning review guidelines to obtain more useful feedback on research programs. GPRA's emphasis on reporting program outcomes was the impetus for changes in NSF's process to include an assessment of how well the results of research programs advance NSF outcome goals. NSF characterizes itself as a learning organization. As such, it applies lessons learned to improving feedback processes in order to keep pace with accountability demands and to obtain more useful information about how completed research contributes to NSF's mission.

Assuring Data Quality

Agencies used two main strategies to meet the demand for better quality data. On their own or with partners, they developed and improved administrative data systems as an aid in obtaining more relevant and reliable data. And when necessary, agencies arranged for special data collection, specifically for research and evaluation use. Initiating new data collection might be warranted by constraints in existing data systems or the excessive cost of modifying those systems.

Improving Administrative Systems

The Coast Guard has developed or improved accounting, financial, and performance reporting systems to enhance access to data on program operations. The Coast Guard, with its diverse program missions (for example, Search and Rescue, Drug Interdiction, and Aids to Navigation) deploys staff and equipment in multiple tasks. The Coast Guard's Abstract of Operations System is the primary source used to identify the allocation of Coast Guard resources and effort. The database tallies the hours spent operating Coast Guard boats and aircraft, allowing the Coast Guard to understand how assets are being used in meeting missions. Managers receive monthly reports and budget officials found this information useful for preparing performance-based budgeting scenarios.

HUD relied on management information systems (MIS), comprised of grantee reports, to keep up with program activities. The data provided critical information on how grant money is being used and what services are received. An official at HUD noted, "Information systems are critical and are becoming more critical every day," but described establishing a

national MIS for CDBG as “excruciating work.” Because of the diversity of CDBG grantees and their activities, it has been difficult to obtain good quality data on a wide range of activities. HUD has improved the quality of information by working with grantees to promote complete and accurate reporting and by automating data collection. With automated data collection, HUD can monitor the completeness of information, edit the data for possible errors, and easily transmit queries arising from those edits back to the source. The CDBG MIS is owned by the program office, which acknowledged the valuable development assistance received from the central analytic office.

HUD officials also noted that, particularly when service delivery rests with a third party, agencies must develop evaluation plans sufficiently in advance to ensure collection of data essential to the evaluation. To evaluate new programs or initiatives, they thought evaluation plans identifying necessary data should be prepared during program development.

Conducting Special Data Collections

Some evaluations rely on data specially collected for that study. For example, agencies may contract out to experienced researchers who collect highly specialized or resource-intensive data. Alternatively, agencies may create specialized data systems. Rather than impose requirements on state program administrative data, NHTSA developed a common data set by extracting standardized data from the states’ systems. NSF developed a special peer review process to obtain data on program outcomes.

The Coast Guard may contract out specialized data collection because a particular research skill is needed or because sufficient staff are not available. For example, the Coast Guard, the U.S. Customs Service, and ONDCP jointly sponsored a study on measuring the deterrent effect of enforcement operations on drug smuggling. To determine how smugglers assess risk and what factors influence their drug smuggling behavior, the study included interviews with high-level cocaine smugglers in federal prisons. This aspect of the study required specialized data collection and interviewing acumen beyond their staff’s expertise. In other drug interdiction and deterrence studies cosponsored with ONDCP, the Coast Guard contracted with the federally sponsored Center for Naval Analyses, which could provide specific services needed for prison interviews and the substantial data collection required.

NHTSA devised a strategy to create a common national data set from varied state data. The Fatality Analysis Reporting System (FARS),

established in 1975, provides detailed annual reports on all fatal motor vehicle crashes during the preceding year, in the 50 states, the District of Columbia, and Puerto Rico. FARS crash record data files contain more than 100 coded data elements characterizing the crash, vehicles, and people involved. Data on crashes must be compiled separately, by state, from multiple source documents (police accident reports and medical service reports) and state administrative records (vehicle registrations and drivers' licenses). NHTSA trains state staff and supervises the coding of the myriad data elements from each state into the common format of standard FARS data collection forms. Training procedures for each state must typically give extensive attention to the detailed content and form of the state systems for compiling police accident reports and other records. These systems often differ between states. Some data items are available from multiple sources within a state, which facilitates cross-checking information accuracy.

NHTSA uses a variety of quality control procedures to assess and ensure the accuracy of several public use data files. The ongoing collection, compilation, and monitoring of these statistical data series greatly facilitates analysis of variation in these data. Such analyses, in turn, lay the foundation for continuing improvements in measurement and in data quality assurance. In addition, the scientific standards that guide NHTSA data quality assurance (1) reflect joint endeavors with other major federal statistical agencies (for example, the Federal Committee on Statistical Methodology) and (2) respond to oversight of federal statistical standards by OMB.¹⁰

To assess research outcomes, NSF created specialized data by using peer review assessments to produce qualitative indicators. To provide credible data to meet GPRA requirements, NSF sought and obtained approval from OMB for the use of nonquantitative performance indicators for assessing outcome goals. Quantitative measures such as literature citations were considered inadequate as an indicator of making substantive scientific contributions. Instead, NSF uses an alternative format—a qualitative assessment of research outcomes—relying on the professional judgment of peer reviewers to characterize their programs' success in making

¹⁰See *The Department of Transportation's Information Dissemination Quality Guidelines* (<http://dmses.dot.gov/submit/dataqualityguidelines.pdf>), as well as the Bureau of Transportation Statistics' *Guide to Good Statistical Practice* (see www.bts.gov).

contributions to science. In order to obtain these new data, questions and criteria were added to the COV review templates.

Obtaining Expertise

The five agencies we reviewed invested in training staff in research and evaluation methods, but frequently relied on outside experts to obtain the specialized expertise needed for evaluation. NHTSA, however, maintains in-house a sizeable staff of analysts skilled in measurement and statistics to develop its statistical series and to identify and evaluate safety issues. In addition, HUD, as well as HHS through ACF and ASPE, supported training for program partners to take prominent roles in evaluating their own programs.

ACF's long-standing collaborative relationship with ASPE helped build the agency's expertise directly—through advising on specific evaluations, as well as indirectly—through building the expertise of the research community that conducts those evaluations. ASPE coordinates and consults on evaluations conducted throughout HHS. ACF staff described getting intellectual support from ASPE—as well as sharing in joint decisions and pooling dollar resources—which boosted the credibility of their work in ACF. At ACF, skills in statistics or research are not enough. They also require people with good communication skills, who can explain the benefits of participation in evaluations to states and localities. For decades, ASPE has funded evaluations, as well as research on poverty, by academic researchers, contract firms, and state agencies. ASPE staff described their investment in poverty research as providing additional assets for evaluation capacity because, in the field of poverty research, the academic world overlaps with the contract firms. They believe this means that (1) better research gets done because prominent economists and sociologists are involved and (2) research on poverty is better integrated with policy analysis than in other fields. For example, agency staff noted that their state agency partners run the National Association for Welfare Research and Statistics, but academics and contractors also participate in National Association conferences. Agency staff also noted that the readability of researchers' reports had improved over time, as researchers gained experience with communicating to policymakers.

The Coast Guard builds capacity in-house and has developed a training program that encourages selected military officers to obtain a Masters in Public Administration (MPA) degree. The Coast Guard selects experts who already have military experience. After receiving a degree, staff are required to do 3- or 4-year payback tours of duty at headquarters, in the role of evaluation analyst, before returning as officers to the field. Staff

trained in operations research might do more statistical analysis at headquarters; those who studied policy and public administration might be more involved in strategic planning and evaluation. The rotations provide (1) field officers with analytic and policy experience and (2) headquarters administrative and planning offices with field experience.

To lay the groundwork for port security planning following the September 11 terrorist attacks, the Coast Guard initiated a process for assessing, over a 3-year period, security conditions of 55 ports. The agency contracted with TRW Systems to conduct detailed vulnerability assessments of these ports. The Coast Guard also contracts for special studies with the agency's Research and Development Center, the Center for Naval Analyses, and the American Bureau of Shipping. In some instances, the Coast Guard used a contractor because the necessary staff were unavailable in-house to collect certain types of data; for example, a national observational study of boaters' use of personal flotation devices (such as life jackets); and a Web-based survey of how mariners use various navigational aids, such as buoys and electronic charting.

NSF, because of the broad array of subject matter disciplines it covers, brings in for a COV, knowledgeable experts from the scientific and engineering communities. COV reviewers must be familiar with their research areas to be able to assess the contribution of funded research to NSF's goals of supporting cutting-edge science. As an approach, peer review involves dozens of outside experts and can be costly; however, because selection confers prestige, researchers are willing to donate their time to the agency. NSF strives to protect COV independence by excluding researchers who are current recipients of NSF awards. In addition, to examine broader issues than a particular research program, NSF may contract with the National Academy of Sciences or the National Institutes of Health for a special study. For other issues that pertain to changes in a field of research or the need for a new strategic direction for research, NSF may put together a blue ribbon panel of experts to provide advice, direction, and guidance.

Providing Technical Expertise to Program Partners

Because of their reliance on state and local agencies for both implementing and evaluating their programs, some of the reviewed agencies found it necessary, in order to improve data quality, to help develop state and local evaluation expertise. In HHS, ACF and ASPE have used several strategies to help develop such expertise. ASPE provided states and counties with grants to study applicants, caseload dynamics, and those who leave welfare. Because states sometimes play a major role in collecting and analyzing data for evaluations, ASPE supported reports

and conferences on data collection and analysis methods, for example, on linking administrative data and research uses of administrative data.

Beginning in 1998, ACF has sponsored annual Welfare Reform Evaluation conferences that bring together state evaluation and policy staff, researchers, and evaluators to share findings and improve the quality and usefulness of welfare reform evaluation efforts. To help develop the next generation of welfare experiments, and engage some states that had not previously been involved, ACF provided planning grants and technical assistance. With the help of a contractor, ACF met with state officials to examine the lessons learned from previous state experiments and help them design their own.

HUD also provides technical assistance to assist local program partners design and manage their programs. HUD provides funding to strengthen the capabilities of program recipients or providers—typically housing or community development organizations. HUD also provides extensive training in monitoring project grants and encourages risk-based monitoring and the flagging of potential problems. A trustworthy administrative database is critical and provides HUD with the information it needs for oversight of how funds are being used.

Building Collaborative Partnerships

The five agencies used collaborative partnerships to obtain access to needed data and expertise for evaluations. Several of these collaborative partnerships developed in pursuit of common goals. Whereas program structures, such as state grants, may create program partners, it often took time and effort to develop collaborative partners. To accomplish the latter, some agencies actively educated program partners and stakeholders about evaluations and solicited their involvement.

Engaging state program partners in evaluation can be difficult, given (1) the voluntary nature of evaluation of state welfare-to-work demonstrations since the waiver evaluation requirement was removed in the 1996 reforms and (2) the risks and burdens of following research protocols. In addition, states may have new ethical reservations—since the 1996 reforms put a time limit on families' receipt of benefits—about withholding potentially helpful services. ACF must therefore entice states to be partners in evaluations that require random assignment. One strategy is to provide funding for the evaluation: ACF used to share funding with the states 50-50. Another is to explain the benefit to them of obtaining rigorous feedback on how well their program is working. ACF also relies on a history of credible and reliable research. To help gain the cooperation

of state and local officials, the agency can point to the good federal-state cooperation it has developed in numerous locations, and show that random assignment is practical.

The poverty research community has not only provided expertise for the state welfare evaluations but also helped build congressional support for those evaluations. For example, researchers briefed congressional committees on evaluation findings, as well as the power of experimental research to reliably detect program effects. The involvement of researchers who are prominent economists and sociologists also helped in drawing lessons from individual evaluations into a cumulative policy-relevant knowledge base. This interconnected web of diverse stakeholders interested in welfare reform—the researchers, the agency, the states, and Congress—has sustained and strengthened a program of research that uses evaluation findings for both program accountability and improvement.

HUD's PD&R takes advantage of opportunities to involve a greater diversity of perspectives, methods, and researchers in HUD research by forming active partnerships with researchers, as well as practitioners, advocates, industry groups, and foundations. A notable illustration is HUD's involvement with the Aspen Institute's Roundtable on Comprehensive Community Initiatives for Children and Families.¹¹ The Roundtable, established in 1992, is a forum for groups engaged in these initiatives to discuss challenges and lessons learned. In 1994, the Roundtable formed the Steering Committee on Evaluation to address key theory and methods challenges in evaluating community initiatives. Along with funding from 11 foundations to support the Roundtable, specific grant funds were provided by the Annie E. Casey Foundation, the Ford Foundation, HUD, HHS, and Pew Charitable Trusts. To ensure that causal links and the role of context are fully understood, the Steering Committee sponsored projects to, for example, clarify and determine outcome indicators and identify methods for collecting and analyzing data.

¹¹Comprehensive Community Initiatives are neighborhood-based efforts to improve the lives of individuals and families in distressed neighborhoods by working comprehensively across social, economic, and physical sectors. The Roundtable, a forum for addressing challenges and lessons learned, now includes about 30 foundation sponsors, program directors, technical assistance providers, evaluators, and public sector officials.

Factors That Impede Building Evaluation Capacity

Although agencies used a variety of strategies to maximize evaluation capacity, they also cited factors that impede conducting evaluations or improving evaluation capacity, including the following:

- Constraints on spending program resources on oversight: Some agency officials claimed that the lack of a statutory mandate or dedicated funds for evaluation impeded investing program funds to conduct studies or to improve administrative data.
- Local control over the design and implementation of flexible programs: To meet local needs, the discretion given to state and local agencies in many federal programs can make it difficult to set federal goals and describe national results. Moreover, variation in evaluation capacity at the local level can impede the collection of uniform, quality data on program performance. As one official noted, when data are derived from data systems built by states to serve their own needs, federal agencies should expect to pay to get data consistency across states.
- Restrictions on federal information collection: Some agency officials voiced concerns about OMB's reviews of agencies' proposed data collection per the Paperwork Reduction Act. They claimed that these reviews constrained their use of some standard research procedures, such as extensively pilot-testing surveys. They also claimed that the length (up to 4 months) and detailed nature of these reviews impeded the timely acquisition of information on program performance.

Observations

The five agencies we reviewed employed various strategies to obtain useful evaluations of program effectiveness. Just as the programs differed from one another, so did the look and content of the evaluations and so did the types of challenges faced by agencies. As other agencies aim to develop evaluation capacity, the examples in this report may help them identify ways to obtain the data and expertise needed to produce useful and credible information on results.

Whether evaluation activities were an intrinsic part of the agency's history or a response to new external forces, learning from evaluation allowed for continuous improvements in operations and programs, and the advancement of a knowledge base. In addition, each agency tied evaluation efforts to accountability demands fostered by GPRA.

Because identifying opportunities for program improvement was so important in sustaining management support for evaluation in these five agencies, other agencies may be more likely to support and use the results of evaluations that are designed to explain program performance than

those that focus solely on whether results were achieved. Similarly, OMB's PART reviews might be useful in encouraging agencies to conduct and use evaluations if budget discussions are focused on what agencies have learned from evaluations about how to improve performance.

Many, if not most, federal agencies rely on third party efforts to help them achieve goals. Agencies might benefit from the examples we present of agencies actively educating and involving program partners as a way to leverage resources and expertise and meet their partners' needs as well.

Agency Comments

HSS and HUD provided technical comments that were incorporated where appropriate throughout the report. HUD pointed out that advance planning was required to ensure collection of key data for an evaluation. We included this point in the discussion of assuring data quality.

We are sending copies of this report to relevant congressional committees and other interested parties. We will also make copies available on request. In addition, the report will be available at no charge on the GAO Web site at <http://www.gao.gov>.

If you have questions concerning this report, please call me or Stephanie Shipman at (202) 512-2700. Valerie Caracelli also made key contributions to this report.



Nancy Kingsbury
Managing Director, Applied Research and Methods

Bibliography

Boyle, Richard, and Donald Lemaire (eds.) *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick, N.J.: Transaction Publishers, 1999.

Committee on Science, Engineering, and Public Policy; National Academy of Sciences; National Academy of Engineering; and Institute of Medicine. *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. Washington, D.C.: National Academy Press, 1999.

Compton, Donald W., Michael Baizerman, and Stacey Hueftle Stockdill (eds.). "The Art, Craft, and Science of Evaluation Capacity Building." *New Directions for Evaluation* 93 (spring 2002).

Fulbright-Anderson, Karen, Anne C. Kubisch, and James P. Connell (eds.). *New Approaches to Evaluating Community Initiatives. Vol. 2: Theory, Measurement, and Analysis*. Washington, D.C.: Aspen Institute Roundtable on Comprehensive Community Initiatives for Children and Families, 1998.

Gueron, Judith M. "Presidential Address—Fostering Research Excellence and Impacting Policy and Practice: The Welfare Reform Story." *The Journal of Policy Analysis and Management*, 22, no. 2 (spring 2003): 163-74.

Gueron, Judith M., and Edward Pauly. *From Welfare to Work*. New York: Russell Sage Foundation, 1991.

Newcomer, Kathryn E., and Mary Ann Scheirer. "Using Evaluation to Support Performance Management: A Guide for Federal Executives." The PricewaterhouseCoopers Endowment for the Business of Government, Innovations Management Series (January 2001).

Office of Management and Budget. "Assessing Program Performance for the FY 2004 Budget."
http://www.whitehouse.gov/omb/budintegration/part_assessing2004.html (April 2003).

Office of Management and Budget. "Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports." Circular no. A-11, pt. 6. (June 2002).

Office of Management and Budget. "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies." *Federal Register* 67, no. 36 (February 22, 2002).

Office of Management and Budget. *Measuring and Reporting Sources of Error in Surveys*. Statistical Policy Working Paper 31, July 2001. <http://www.fcsm.gov/reports#fcsm>. (April 2003).

Office of Management and Budget. *Performance and Management Assessments, Budget of the United States Government, Fiscal Year 2004*. Washington, D.C.: U.S. Government Printing Office. <http://www.whitehouse.gov/omb/budget/fy2004> (April 2003).

Office of Management and Budget. *The President's Management Agenda, Fiscal Year 2002*. http://www.whitehouse.gov/omb/budintegration/pma_index.html (April 2003).

Office of National Drug Control Policy. *Measuring the Deterrent Effect of Enforcement Operations on Drug Smuggling, 1991-1999*. Prepared by Abt Associates, Inc. Washington, D.C.: August 2001. <http://www.whitehousedrugpolicy.gov/publications> (April 2003).

Rossi, Peter H., and Katharine C. Lyall. *Reforming Public Welfare: A Critique of the Negative Income Tax Experiment*. New York: Russell Sage Foundation, 1976.

Sonnichsen, Richard C. *High-Impact Internal Evaluation: A Practitioner's Guide to Evaluating and Consulting Inside Organizations*. Thousand Oaks, Calif.: Sage Publications, 1999.

U.S. Department of Transportation. *The Department of Transportation's Information Dissemination Quality Guidelines*. October 1, 2002. <http://www.bts.gov/statpol> (April 2003).

U.S. Department of Transportation. Bureau of Transportation Statistics. *BTS Guide to Good Statistical Practice*. September 2002. (<http://www.bts.gov/statpol/guide/index.html>) (April 2003).

Related GAO Products

Welfare Reform: Job Access Program Improves Local Service Coordination, but Evaluation Should Be Completed. [GAO-03-204](#). Washington, D.C.: December 6, 2002.

Coast Guard: Strategy Needed for Setting and Monitoring Levels of Effort for All Missions. [GAO-03-155](#). Washington, D.C.: November 12, 2002.

HUD Management: Impact Measurement Needed for Technical Assistance. [GAO-03-12](#). Washington, D.C.: October 25, 2002.

Program Evaluation: Strategies for Assessing How Information Dissemination Contributes to Agency Goals. [GAO-02-923](#). Washington, D.C.: September 30, 2002.

Performance Budgeting: Opportunities and Challenges. [GAO-02-1106T](#). Washington, D.C.: September 19, 2002.

Surface and Maritime Transportation: Developing Strategies for Enhancing Mobility: A National Challenge. [GAO-02-775](#). Washington, D.C.: August 30, 2002.

Port Security: Nation Faces Formidable Challenges in Making New Initiatives Successful. [GAO-02-993T](#). Washington, D.C.: August 5, 2002.

Public Housing: New Assessment System Holds Potential for Evaluating Performance. [GAO-02-282](#). Washington, D.C.: March 15, 2002.

National Science Foundation: Status of Achieving Key Outcomes and Addressing Major Management Challenges. [GAO-01-758](#). Washington, D.C.: June 15, 2001.

Motor Vehicle Safety: NHTSA's Ability to Detect and Recall Defective Replacement Crash Parts Is Limited. [GAO-01-225](#). Washington, D.C.: January 31, 2001.

Program Evaluation: Studies Helped Agencies Measure or Explain Program Performance. [GAO/GGD-00-204](#). Washington, D.C.: September 29, 2000.

Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information. [GAO/GGD-99-139](#). Washington, D.C.: July 30, 1999.

Federal Research: Peer Review Practices at Federal Science Agencies Vary. [GAO/RCED-99-99](#). Washington, D.C.: March 17, 1999.

Managing for Results: Measuring Program Results That Are Under Limited Federal Control. [GAO/GGD-99-16](#). Washington, D.C.: December 11, 1998.

Grant Programs: Design Features Shape Flexibility, Accountability, and Performance Information. [GAO/GGD-98-137](#). Washington, D.C.: June 22, 1998.

Program Evaluation: Agencies Challenged by New Demand for Information on Program Results. [GAO/GGD-98-53](#). Washington, D.C.: April 24, 1998.

Program Measurement and Evaluation: Definitions and Relationships [GAO/GGD-98-26](#) Washington, D.C.: April, 1998.

Measuring Performance: Strengths and Limitations of Research Indicators. [GAO/RCED-97-91](#). Washington, D.C.: March 21, 1997.

Program Evaluation: Improving the Flow of Information to the Congress. [GAO/PEMD-95-1](#). Washington, D.C.: January 30, 1995.

GAO's Mission

The General Accounting Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily E-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are \$2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
 TDD: (202) 512-2537
 Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm

E-mail: fraudnet@gao.gov

Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, managing director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548