United States General Accounting Office

GAO

Report to Congressional Committees

May 1997

MANAGING FOR RESULTS

Analytic Challenges in Measuring Performance

GAO/HEHS/GGD-97-138

GAO

United States
General Accounting Office
Washington, D.C. 20548

Health, Education, and
Human Services Division

B-276736

May 30, 1997

The Honorable Fred Thompson
Chairman
The Honorable John Glenn
Ranking Minority Member
Committee on Governmental Affairs
United States Senate

The Honorable Dan Burton
Chairman
The Honorable Henry A. Waxman
Ranking Minority Member
Committee on Government Reform and Oversight
House of Representatives

Seeking to promote improved government performance and greater public confidence in government through better planning and reporting of the results of federal programs, the Congress enacted the Government Performance and Results Act of 1993 (GPRA), which is referred to as "the Results Act" and "GPRA." The Act established a governmentwide requirement for agencies to identify agency and program goals and to report on their results in achieving those goals. Recognizing that few programs at the time were prepared to track progress toward their goals, the Act specifies a 7-year implementation time period and requires the Office of Management and Budget (OMB) to select pilot tests to help agencies develop experience with the Act's processes and concepts. The Results Act includes a pilot phase during which about 70 programs, ranging from the U.S. Geological Survey's National Water Quality Assessment Program to the entire Social Security Administration, were designated as GPRA pilot projects. These and other programs throughout the major agencies have been gaining experience with the Act's requirements. GPRA mandates that we review the implementation of the Act's requirements in this pilot phase and comment on the prospects for compliance by federal agencies as governmentwide implementation begins in 1997. This report is one component of our response to that mandate. Specifically, this report answers the following questions: (1) What analytic and technical challenges are agencies experiencing as they try to measure program performance? (2) What approaches have they taken to address these challenges? And, in particular, because program evaluation studies are similarly focused on measuring progress toward program goals and objectives, (3) How have agencies made use of program evaluations or

evaluation expertise in implementing performance measurement? Indeed, the Act recognizes and encourages a complementary role for program evaluation by requiring agencies to describe its use in performance planning and reporting.

To obtain this information, we conducted structured interviews with program officials in 20 departments and major agencies with experience in performance measurement. Generally, in each agency, we selected one official GPRA pilot program and one other program that had begun to measure program performance. We selected programs to represent diversity in program purpose, size, and other factors that we thought might affect their experience. For each program, we attempted to interview both the program official responsible for performance measures and a program evaluator or other analyst who had assisted in this effort. Since no evaluator was identified in some programs, while in others, the evaluator was the person responsible for the performance measurement effort, we conducted 68 structured interviews with officials from 40 programs. We asked program officials to rate the difficulty of challenges or tasks at each of four stages in the performance measurement process that we defined for the purposes of this review:

- identifying goals: specifying long-term strategic goals and annual performance goals that include the outcomes of program activities;
- developing performance measures: selecting measures to assess programs' progress in achieving their goals or intended outcomes;
- collecting data: planning and implementing the collection and validation of data on the performance measures; and
- analyzing data and reporting results: comparing program performance data with the annual performance goals and reporting the results to agency and congressional decisionmakers.

Then, for each stage, we asked program officials to describe how they approached their most difficult challenge and whether and how they used prior studies and technical staff. A more complete description of the scope of this review is included in appendix I.

Results in Brief

The programs included in our review encountered a wide range of serious challenges—93 percent of the officials we surveyed reported at least one as a great or very great challenge. In addition, some were not very far along in implementing the steps required by the Results Act. Eight of the 10 tasks rated most challenging emerged in the two relatively early stages

of the performance measurement process: identifying goals and developing performance measures. For example, in the stage of identifying goals, respondents found it particularly difficult to translate long-term strategic goals into annual performance goals. This was often because the program had a long-term mission that made it difficult to predict the level of results that might be achieved on an annual basis.

In developing both goals and performance measures, respondents found it difficult to move beyond a summary of their program's activities—such as the number of clients served—to distinguish the desired outcome or result of those activities—such as the improved health of the individuals served or the community at large. For some, the concept of "outcome" was unfamiliar and difficult especially for program officials focused on day-to-day activities. Sometimes selecting an outcome measure was impeded, instead, by conflicting stakeholder views of the program's intended results or by anticipated data collection problems. Issues in the data collection stage were rated as less serious and revolved around the programs' lack of control over data that third parties collected, but programs may have avoided some data issues through selection of measures for which data already existed.

The greatest challenge in the analysis and reporting stage was separating a program's impact on its objectives from the impact of external factors, primarily because many federal programs' objectives are the result of complex systems or phenomena outside the program's control. In such cases, it is particularly challenging for agencies to confidently attribute changes in outcomes to their program—the central task of program impact evaluation. Although the Act does not require impact evaluations, it does require programs to measure progress toward achieving their goals and explain why a performance goal was not met. Because they recognized that simple examination of outcome measures would not accurately reflect their program's performance, many of the respondents believed that they ought to separate the influence of other factors on their program's goals in order to establish program impact.

The programs we reviewed had applied a range of analytic and other strategies to address these challenges. To overcome uncertainties in formulating performance goals that were achievable on an annual basis, some programs had adopted a multiyear planning horizon for their performance goals, while others had modified their annual goals to target more proximate ones over which they had more control. A wide variety of approaches was used to help define performance measures, including

developing a model of the relationships between federal, state, and local government activities to identify the uniquely federal role. Programs that found reliance on others' data as their greatest data collection challenge tended to either introduce data verification procedures or search for alternative data sources. The programs employed several different approaches to attempt to isolate a program's impact from other influences, including conducting special studies and monitoring external factors at the subnational level, where their influence was easier to observe. Overall, the programs we reviewed had somewhat more difficulty in resolving their most difficult challenges related to selecting measures and analyzing performance than in identifying goals and collecting data; they were less likely to have developed an approach to meeting these challenges, and they reported less confidence in the approaches they had developed.

Because they had either volunteered to be GPRA pilots or had already begun implementing performance measurement, the programs included in our review were likely to be better suited or prepared for conducting performance measurement than most federal programs. In addition, they had the advantage of technical resources: half of these programs had been the subject of previous evaluations, and almost all had access to staff trained or experienced in performance measurement or program evaluation. Most of our respondents found this assistance helpful, and many said they could have used more such assistance. For example, an evaluator assisting one program adapted a data collection instrument from a prior study to collect data on outcomes that were considered difficult to measure. Also, an administrator trained in evaluation methods, faced with program outcomes known to be subject to external influences, developed a series of outcome measures and looked at the similarity of results across them to assess program performance.

The challenges experienced by the projects that are pilot testing the Act's requirements suggest that (1) more typical federal programs may find performance measurement to be an even greater challenge, particularly if they do not have access to program evaluation or other technical resources; and (2) full-scale implementation will require several iterations to develop valid, reliable, and useful performance reporting systems. In addition, in cases in which factors outside the program's control are acknowledged to have significant influence on key program results, it may be important to supplement performance measure data with impact evaluation studies to provide an accurate picture of program effectiveness.

Background

The Results Act seeks to improve the efficiency, effectiveness, and public accountability of federal agencies as well as to improve congressional decision-making. It aims to do so by promoting a focus on program results and providing the Congress with more objective information on the achievement of statutory objectives. The Act outlines a series of steps whereby agencies are required to identify their goals, measure performance, and report on the degree to which those goals were met. The Act requires executive branch agencies to develop, by the end of fiscal year 1997, a strategic plan and to submit their first annual performance plan to OMB in the fall of 1997. Starting in March of the year 2000, each agency is to submit a report comparing its performance for the previous fiscal year with the goals in its annual performance plan. However, OMB also asked all agencies to include performance measures, if available, with their budget requests for fiscal year 1998 in order to encourage planning for meeting the Act's requirements. (App. II describes the Act's requirements in more detail.) For the purpose of this review, we identified four stages in the performance measurement process to represent the analytic tasks involved in producing these documents. Figure 1 depicts the correspondence between these stages and the Act's requirements.

Figure 1: A Comparison of Our Four Stages of the Performance Measurement Process With GPRA Requirements

Stage of the Performance Measurement Process	Stage 1: Identifying Goals	Stage 2: Developing Performance Measures	Stage 3: Collecting Data	Stage 4: Analyzing Data and Reporting Results
GPRA Requirement	**Strategic Plan**	**Performance Plan**	**Performance Report**	
	• Identify the agency's mission and long-term, strategic goals • Describe how the agency will achieve the goals through its activities and resources • Describe how the agency's annual performance goals are related to its long-term goals • Identify factors external to the agency that could affect goal achievement • Describe program evaluations used in establishing or revising the goals and include a schedule of future evaluations	• Specify annual performance goals for each program activity • Identify the performance measures the agency will use to assess its progress • Describe how the data will be verified and validated	• Compare performance data for the previous fiscal year with the goals in the annual performance plan • Describe plans for meeting unmet goals or explain why a goal shoud be modified • Summarize findings of program evaluations completed during the fiscal year	

In the past, some agencies have conducted program evaluations to provide information to program managers and the Congress about whether a program is working well or poorly, and why. Most evaluations of program effectiveness, or program impact, include the basic planning and analysis steps that the Act requires agencies to take: defining and clarifying program goals and objectives, developing measures of program outcomes, and collecting and analyzing data to draw conclusions about program results. However, program impact evaluation goes further to establish the causal connection between outcomes and program activities, separate out the influence of extraneous factors, develop explanations for why those outcomes occurred, and thus isolate the program's contribution to those changes. Thus, where programs are expected to produce changes as a

result of program activities, such as job placement activities for welfare recipients, outcome measures can tell whether the welfare caseload decreased. However, a systematic evaluation of a program's impact would be needed to assess how much of the observed change was due to an improved economy or to the program. In addition, a systematic evaluation of how a program was implemented can provide important information about why a program did or did not succeed and suggest ways to improve it. However, because the tasks involved raise technical and logistical challenges, evaluating program impact generally requires a planned study and, frequently, considerable time and expense.

The Results Act recognizes the complementary nature of performance measurement and program evaluation, requiring a description of previous program evaluations used and a schedule for future program evaluations in the strategic plan, and a summary of program evaluation findings in the annual performance report. In addition, because of the similarities between performance measurement and program evaluation, we expected that experience with or access to expertise in program evaluation would assist agencies in addressing the challenges of performance measurement. Therefore, we included in our survey programs other than the official GPRA pilots that were said to have had experience in measuring program results and that may have had program evaluation experience. In addition, we interviewed program officials responsible for performance measurement and program evaluators or other analysts who had assisted in this effort, if available, and we asked whether prior studies or technical staff had been involved in the various performance measurement tasks.

Agencies Are Still in Early Implementation Phase of Performance Measurement

Despite having volunteered to begin measuring program performance, most of the programs we reviewed had not yet gone through all the steps of the performance measurement process. Almost all our respondents (over 96 percent) reported that their programs had begun the first three stages of performance measurement, and 85 percent had started data analysis and reporting. But only about 27 percent had actually completed all four stages (see table 1). Overall, programs were furthest along with the stage of identifying goals, and least with the reporting stage, but they did not, of course, need to "complete" one stage before starting another, because performance measurement is recognized to be an iterative process in which measures will be improved over time. For example, if data are unavailable for the annual performance report, agencies are permitted to provide whatever data are available, with a notation as to their incomplete status, and to provide the data in subsequent reports.

Table 1: Percentage of Respondents Reporting That Their Programs Have Completed Performance Measurement Stages (for the Total Sample and Selected Subgroups)

Program characteristic	Identifying goals	Developing performance measures	Collecting data	Analyzing data and reporting results	Completed at least one round of all four stages
Total sample	66%	57%	54%	53%	27%
Program purpose					
Provide services or military defense	64	59	54	49	26
Develop information	65	65	60	60	37
Administer regulations	78	33	44	56	11
GPRA status					
Official pilot	87	67	60	70	38
Other	50	50	50	40	19
Annual budget					
Less than $100 million	77	62	77	62	42
Between $100 million and $1 billion	59	48	41	48	15
Greater than $1 billion	64	64	50	46	29
Locus of control					
Federal	70	62	50	68	30
State	67	57	52	47	18
Local or quasigovernmental organization	89	56	90	73	36

Regulatory programs were far behind in completing at least one round of all four stages (11 percent), apparently because of their difficulty with specifying performance measures and data collection. Official GPRA pilots were twice as likely to have gone through all four stages as other programs (38 percent and 19 percent, respectively), in part because they were much further along in goal identification than the other programs (87 percent compared with 50 percent). Staff from smaller programs reported their programs were much further along (42 percent had completed all four stages) and were more likely to have completed at least one reporting cycle than larger programs. This could stem partly from the fact that most of the small programs in our sample were GPRA pilots (85 percent). As such, many would have already submitted to OMB both an annual performance plan and an annual performance report. However, the small programs as a whole were also more likely to have completed data collection than the GPRA pilots as a group (77 percent compared with 60 percent). In general, little difference in progress was seen between

state- and federally administered programs across the first three stages, but state-administered programs were not as far along in analysis and reporting, or in completing a full cycle of the process, as programs run at either the federal or local level. Differences in progress among programs with different funding sources were inconsistent.

Programs' Greatest Challenges Generally Came in the Early Stages of Implementing Performance Measurement

Almost all of the programs included in our review encountered serious challenges—93 percent of our respondents rated at least 1 of 30 potential challenges as a great or very great challenge. Most respondents (74 percent) identified a great challenge in the stage of identifying goals; 69 percent identified at least one in the stage of developing performance measures. Fewer reported encountering a great challenge in the later stages of data collection and reporting results (50 and 34 percent, respectively).

To indirectly assess which of our four stages of performance measurement—identifying goals, developing measures, collecting data, or analyzing and reporting results—provided the most difficult challenges for these agencies, we rank-ordered each of 30 potential challenges by respondents' mean ratings of their difficulty. We found 8 of the 10 challenges with the highest mean ratings among the two early, relatively conceptual stages of specifying the program's goals—especially as the outcomes or results of program activities—and selecting objective, quantifiable measures of them (see table 3). Three challenges pertained to the stage of identifying goals and five to developing measures. Issues in the two later stages of data collection and analysis were generally rated less challenging except for two items—ascertaining the accuracy and quality of performance data and separating a program's impact on its objectives from the impact of external factors—which, although not specifically required by the Act, is often needed to confidently attribute results to the program. (In this and subsequent tables, the number of valid cases reflects those that had begun that performance measurement stage and experienced the challenge.)

Table 2: The Performance Measurement Stage and Mean Rating of the 10 Challenges Rated Most Difficult by Respondents

Analytic stage	Challenge	Mean rating[a]	Valid cases
Identifying goals	Translating general, long-term strategic goals to more specific, annual performance goals and objectives	3.36	59
	Distinguishing between outputs and outcomes	3.27	63
	Specifying how the program's operations will produce the desired outputs and outcomes	3.20	61
Developing performance measures	Getting beyond program outputs—that is, summaries of program activities—to develop outcome measures of the results of those activities	3.52	65
	Specifying quantifiable, readily measurable performance indicators	3.25	65
	Developing interim or alternative measures for program effects that may not show up for several years	3.09	54
	Estimating a reasonable level for expected performance	3.03	60
	Defining common, national performance measures for decentralized programs	2.96	46
Collecting data	Ascertaining the accuracy of and quality of performance data	2.92	60
Analyzing data and reporting results	Separating the impact of the program from the impact of other factors external to it	3.11	45

[a]On a scale of 1 ("little or no challenge") to 5 ("a very great challenge").

In most programs, respondents rated the same general mix of problems as their most difficult, except for the regulatory programs, for which three of their five greatest challenges came from the later two stages. The problem these regulatory programs ranked as most difficult was separating the impact of the program on its objectives from the impact of external factors. They also reported difficulty with ascertaining the accuracy and quality of performance data and with acquiring the exact data wanted and in the form desired. This might be explained by these programs' reliance on the regulated parties themselves to provide data on their own level of compliance.

Across all stages, the official pilots rated the potential challenges we posed as less difficult, on the average, than did the other programs. Pilots also included two challenges from later stages among their top five most difficult—separating the impact of the program from that of external factors and using data collected by others—while the other programs did not. We do not know whether this may have been influenced by the pilots' greater experience than the other programs with a full reporting cycle.

Long-Term Missions, Rare Events, and Difficulties in Conceptualizing Outcomes Made Specifying Annual Goals Difficult

Considering first the challenges in the stage of identifying goals, the three greatest challenges were (1) translating general, long-term strategic goals to more specific, annual performance goals and objectives; (2) distinguishing between outputs and outcomes; and (3) specifying how the programs' operations would produce the desired outputs and outcomes (see table 3).[1] About twice as many respondents rated these as great or very great challenges compared to reducing the program to a few broad, general goals.

[1]We ranked the challenges by their means, by the percentage reporting that they were a great or very great challenge, and by how often each challenge was reported as the greatest challenge encountered in that stage. These different methods resulted for the most part in similar rankings.

Table 3: Respondents' Ratings of the Level of Difficulty Posed by Potential Challenges in Identifying Goals

Potential challenge	Actual extent of challenge: Percentage rating this as a great or a very great challenge	Actual extent of challenge: Mean challenge rating[a]	Actual extent of challenge: Valid cases
Translating general, long-term strategic goals to more specific, annual performance goals and objectives	49	3.36	59
Distinguishing between outputs and outcomes	46	3.27	63
Specifying how the program's operations will produce the desired outputs and outcomes	44	3.20	61
Reconciling potentially conflicting goals	25	2.40	60
Reducing the program to a few broad, general goals	23	2.74	62
Accommodating state or local goals and objectives	18	2.79	38
Identifying critical external factors	19	2.48	58
Specifying objectives for the entire program rather than just certain parts of it	15	2.30	53
Distinguishing this program's goals from those of related programs	13	2.14	56

[a]On a scale of 1 ("little or no challenge") to 5 ("a very great challenge").

In identifying goals (and performance measures), respondents found it difficult to respond to the Act's encouragement for agencies to move beyond summarizing their program's activities—such as measuring the number of clients served— to distinguishing the desired outcome or result of those activities—such as improving the health of the individuals served or the community at large. Some of our respondents explained that translating strategic goals for long-term missions—such as supporting basic science—into annual goals was particularly difficult because annual goals tend to be artificial and hard to analyze given the unpredictable nature of scientific progress. Others reported that the constantly changing nature of their target—for example, a developing business sector or newly democratizing country—made annual, linear progress unlikely. There were also managerial, process issues cited. As one respondent said, "It is easier to get agreement on long-term goals, but once you begin to break them

down into annual objectives and specify how you will achieve them, you get into disagreement over priorities, approaches, and roles."[2]

Distinguishing between outputs and outcomes was found to be a challenge for several reasons. First, some struggled with the basic meaning of the concept of outcome. One respondent noted that OMB's definition of "outcome" varied from one set of guidance to the next. Another reported that the program's administrators still believed that regulations were the outcomes and that whatever happened after a new regulation was issued was beyond their control. Different administrators, staff, and stakeholders defined outcomes in multiple ways and by their regional or national context.

Second, some argued that the nature of their missions made it hard to develop a measurable outcome. For example, when the goal was to prevent a rare event, such as a flood or presidential assassination attempt, the fact that it did not occur is hard to attribute to a particular function. Similarly, some outcomes, like battles won, may not be observed in a given year. Thus, it may be conceptually more difficult to define outcomes for prevention, deterrence, and other programs that respond to rare events.

Third, in addition to conceptual challenges, there were administrative obstacles. One respondent reported that because several states had been developing their own outcome measures for their program for some time, they had sunk costs in their existing information systems. Thus, they were opposed to standardizing the measures solely so that federal administrators could come up with a new, common measure.

Respondents who said that their most difficult problem in identifying goals was specifying how program operations would produce outputs and outcomes did not report anything inherently difficult in building logic models for programs. Rather, they cited many of the other potential challenges as factors that impeded this planning step, such as the role of external factors, the unpredictability of prevention outcomes or outcomes that may take many years to develop, and their lack of leverage over state approaches.

[2]OMB also found, in reviewing agency progress in strategic planning, that virtually every agency had difficulty linking long-range strategic mission and goals with annual performance goals. (John A. Koskinen, OMB, letter to the Honorable Dan Glickman, Secretary of Agriculture, Aug. 9, 1996.)

A Short-Term Focus, Multiple Stakeholders, and Data Constraints Made Specifying Performance Measures Difficult

The challenges rated most difficult, on average, in specifying performance measures were (1) getting beyond program outputs (that is, summaries of program activities) to develop measures of outcomes or the results of those activities; (2) specifying quantifiable, readily measurable performance indicators; and (3) developing interim or alternative measures for program effects that may not show up for several years (see table 4). Similar reasons were given for why each of these challenges was particularly difficult.

Table 4: Respondents' Ratings of the Level of Difficulty Posed by Potential Challenges in Developing Performance Measures

	Actual extent of challenge		
Potential challenge	Percentage rating this as a great or very great challenge	Mean challenge rating[a]	Valid cases
Getting beyond program outputs, that is, summaries of program activities, to develop outcome measures of the results of those activities	49	3.52	65
Specifying quantifiable, readily measurable performance indicators	42	3.25	65
Defining common, national performance measures for decentralized programs	39	2.96	46
Developing interim or alternative measures for program effects that may not show up for several years	37	3.09	54
Estimating a reasonable level for expected program performance	32	3.03	60
Developing qualitative measures such as narrative descriptions where numerical measures could not be had	29	2.84	49
Planning how to compare actual program results with the performance goals	20	2.40	60

[b]On a scale of 1 ("little or no challenge") to 5 ("a very great challenge").

Respondents found that, at the most basic level, defining the specific outcomes desired for their program was difficult to accomplish, but it was also complicated by program-specific conditions. Some said that defining outcome measures required administrators to change from thinking on a day-to-day basis to taking a long-term perspective on what they wanted to accomplish, as indeed the Act intended them to do. Shifting to a long-term

perspective led them to broaden their horizons to consider outcomes over which they rarely have complete control, introducing additional uncertainty. More generally, some respondents observed that "outcome" seemed to be a fuzzier concept than "output," difficult to think through and specify precisely. These tasks were said to be particularly difficult in a volatile, complex policy environment.

In addition, to arrive at an outcome definition that would be broadly accepted, program officials reported having to do a lot of consensus building with stakeholders who often disagreed on the validity of outcome measures. Some reported difficulty in getting state program administrators and other federal stakeholders not only to think beyond their own program operations, as previously noted, but also to conceptualize how those diverse activities were related to a common outcome for the nation as a whole. Others noted that efforts to agree on measures had to overcome program officials' reluctance to be measured except in the most favorable light, concerned, perhaps, with the potential use of performance data to blame program officials rather than improve program functioning.

For others, selecting outcome measures was difficult because it was intertwined with anticipated data collection problems. They noted that a focus on outcomes involves developing new measures, new databases, and, often, learning new measurement techniques. Moreover, the annual reporting requirement was said to force certain issues: for example, annual data collection needs to be orchestrated and routinized, thus either raising additional logistics questions or limiting program officials' choice of measures, if new data collection was not a practical option.

Respondents Blamed the Need to Rely on Others for Their Greatest Data Collection Challenges

Although, in general, the potential challenges in data collection were not considered as difficult as those in other stages, about one-third of our respondents reported that the following were particularly challenging: (1) using data collected by others, (2) ascertaining the accuracy and quality of performance data, and (3) acquiring the data in a timely way (see table 5). However, these programs may have avoided some of the data issues we posed through decisions made in the previous stage to select measures for which the respondents had existing data. Our respondents said that using data collected by others was challenging because it was difficult to ascertain their quality or to ensure their completeness and comparability. The respondents also found a management challenge in attempting to overcome resistance by external data providers to spending money on additional data collection and to sharing costly data. Two

respondents also reported having to deal with deliberate misreporting by other agencies that were trying to justify higher funding levels.

Table 5: Respondents' Ratings of the Level of Difficulty Posed by Potential Challenges in Data Collection

	Actual extent of challenge		
Potential challenge	Percentage rating this as a great or very great challenge	Mean challenge rating[a]	Valid cases
Using data collected by others	33	2.74	46
Ascertaining the accuracy of and quality of performance data	30	2.92	60
Acquiring the data in a timely way	28	2.72	61
Acquiring the exact data wanted and in the form desired	26	2.74	62
Obtaining baseline data for comparison	25	2.69	59
Ascertaining the accuracy of and quality of baseline data	22	2.81	59
Identifying and locating sources of data for the performance measures	11	2.25	63

[a]On a scale of 1 ("little or no challenge") to 5 ("a very great challenge").

The fact that their data were largely collected by others was the most frequent explanation of why ascertaining the accuracy and quality of performance data was a problem. One respondent said that collecting federal data is not a high priority for most states, and thus they do not emphasize the data's accuracy. Documentation of data quality was reportedly often not available or was incomplete. For example, one respondent said that in his area, most state record-keeping is manual and hard to audit. Acquiring the data in a timely way was reported as hindered by lack of adequate database systems; more often it was said to be hindered by a mismatch between the data collection time lines and the reporting cycle.

The Influence of Factors Beyond the Program's Control Makes Attributing the Results to the Program Difficult

When it came to analyzing and reporting performance, one challenge stood out clearly as the most difficult: separating the impact of the program from the impact of other factors external to the program (see table 6). Forty-four percent of respondents who had begun this stage claimed that it was a great or very great challenge. The difficulty was primarily the fact that the outcomes of many federal programs are the result of the interplay of several factors, and only some of these are within the program's control.

Even simple, two-variable interactions are potentially difficult. For instance, if a new weapon system is introduced late in the fleet training cycle, lower-than-expected levels of performance could be caused by problems in the weapon system or in the training program.

Table 6: Respondents' Ratings of the Level of Difficulty Posed by Potential Challenges in Analysis and Reporting

	Actual extent of challenge		
Potential challenge	**Percentage rating this as a great or very great challenge**	**Mean challenge rating[a]**	**Valid cases**
Separating the impact of the program from the impact of other factors external to the program	44	3.11	45
Calculating the outputs and outcomes for any program components	24	2.43	49
Having to modify or develop additional indicators	23	2.60	43
Understanding the reasons for unmet goals or unanticipated results	16	2.25	44
Comparing actual program performance results with the performance goals	13	1.98	47
Translating the results into recommendations for future program improvement and better performance measurement	12	2.24	42
Data that turned out to be inadequate for the intended analysis	11	2.11	44

[a]On a scale of 1 ("little or no challenge") to 5 ("a very great challenge").

More importantly, many programs consist of efforts to influence highly complex systems or phenomena outside government control. In such cases, one cannot confidently attribute a causal connection between the program and its outcomes. Respondents noted that controlling for all external factors in order to measure a program's effect is very difficult in programs that attempt to intervene in highly complex systems such as ecosystems, year-to-year weather, or the global economy. Additionally, respondents pointed to other factors that can exacerbate this problem, such as very long-term outcomes that are difficult to link directly to program activity.

Although the Act does not require agencies to conduct formal impact evaluations, it does require them to (1) measure progress toward achieving their goals, (2) identify which external factors might affect such progress, and (3) explain why a goal was not met. Although few respondents reported difficulty identifying these external factors during the goal identification stage (19 percent, as shown in table 3), actually isolating their impact on the outcomes during analysis was reported to be a more formidable challenge. This could be due either to analytic or to conceptual problems in controlling for the influence of other factors. Nevertheless, because they realized that a simple examination of the outcome measures would not accurately reflect their program's performance, many of our respondents believed that they ought to go to the next step and separate the influence of other factors on their program's goals, in order to establish their program's impact.

Programs Took Varied Approaches to Address Their Most Difficult Challenges

Respondents reported active efforts to address those challenges they identified as most difficult in each of the four stages. The approaches they described covered a range of strategies, from participatory activities (such as consulting with stakeholders or providing program managers with training in reporting outcome data) to applying statistical and measurement methods (such as conducting a customer survey or developing multiple measures of associated program outcomes for an outcome that was difficult to measure directly). Programs applied similar participatory strategies throughout the performance measurement stages but tended to tailor the analytic strategies to the particular challenge, sometimes using quite different approaches to the same challenge. The scope and ingenuity of some of these approaches demonstrate serious engagement in the analytic dimension of performance measurement.

Program officials reported relatively high levels of technical staff involvement across the four performance measurement stages (72 to 82 percent of all those who identified a challenge in those stages; see table 7). Nevertheless, they appeared to have somewhat more difficulty resolving their most difficult challenges in the stages of developing performance measures and analyzing data and reporting results than in the other two stages. Program respondents were more likely to report in these stages (11 and 12 percent, respectively) that their performance measurement team was still trying to determine what to do. Moreover, respondents also reported feeling more successful in their responses to the most difficult challenges in identifying goals and collecting data than with those in selecting measures and in analysis and reporting. This

pattern of experiencing greater satisfaction in their approaches to the challenges in the goal identification and data collection stages was even more apparent when we looked at the single challenge in each stage that the greatest number of respondents considered most difficult.[3]

Table 7: Respondents' Use of Evaluation Resources, Development of Approaches, and Views of Success

	Performance measurement stage			
Item	**Identifying goals**	**Developing performance measures**	**Collecting data**	**Analyzing data and reporting results**
Evaluation resources				
Number of respondents who identified one challenge in the stage as most difficult	61	62	58	42
Percentage who had access to prior studies	82%	81%	84%	87%
Percentage of those who considered prior studies helpful	77%	80%	80%	74%
Percentage who were assisted by technical staff in this stage	72%	82%	81%	74%
Approaches				
Developed[a]	93%	89%	98%	88%
Yet to be developed	7%	11%	2%	12%
Views of success				
Minimally successful	5%	16%	10%	14%
Somewhat successful	7%	22%	16%	14%
Moderately successful	42%	30%	29%	32%
Mostly successful	18%	24%	28%	34%
Very successful	28%	8%	17%	7%

[a]Percentage of approaches to the most difficult challenge in a stage reported by respondents who had identified one challenge as most difficult.

Approaches to Translating Long-Term Goals Into Annual Goals

In the first stage, identifying goals, the challenge respondents most frequently identified as their most difficult was translating the long-term goals established in their strategic plan into annual performance goals. All 12 respondents selecting this challenge as their most difficult (representing 10 programs) reported having developed an approach to this

[3]We did not independently assess the approaches respondents described.

challenge, and most were well satisfied with how it met the challenge.[4] Half rated their approach as mostly to very successful, and half rated it as moderately successful in responding to the challenge. (App. III provides data on respondents' views of the approach they developed and their use of evaluation resources for those who selected this as the most serious challenge in this stage.) This group of respondents was a little less likely than the full sample to report having access to prior studies to develop their approaches to identifying goals. Three-quarters had prior studies to draw on, and three-quarters were assisted by technical staff. All those with access to prior studies generally found them to be helpful.

To address the challenge of specifying annual goals that were consistent with their long-range goals, the respondents reported that they tended either to use other than an annual time period for reporting or to modify the global outcome toward which the goals were directed. (Table 8 shows the types of approaches the programs developed for this challenge and for the second most frequently identified challenge.) For example, two respondents reported that their programs found that setting annual goals was not feasible because of the exploratory and long-range nature of their work. One respondent compared the program's role with that of an investment broker with a portfolio, for which long-term goals are fairly well identified but for which annual expectations are much less certain. He added that because the program operates through the grant-funding mechanism, which is less directive than other forms of financial assistance, it requires an investment perspective. The manager of the second program pointed out that it is difficult to set annual goals for a program targeted on a rapidly changing industry. Both of these programs had adopted a multiyear planning horizon for their performance goals.

[4]Among programs represented by two respondents, in some cases, both identified the same challenge as most difficult. However, in other cases, each respondent identified a different challenge as most difficult.

Table 8: Approaches Taken to the Most Difficult Challenges in Identifying Goals

Challenge	Number of respondents[a]	Approach to identifying goals
Translating long-term goals into annual performance goals	12	Specified performance goals over an extended period
		Focused annual goals on proximate outcomes
		Developed a conceptual model to specify annual goals
		Focused annual goals on short-term strategies for achieving long-term goals
		Developed a qualitative approach
		Involved stakeholders
Distinguishing between outputs and outcomes	9	Clarified definitions of output and outcome
		Focused on known, quantifiable outcomes
		Focused on projected outputs
		Surveyed customers to identify outcomes
		Involved stakeholders

[a]Number of respondents who identified the challenge as most difficult and had developed an approach to that challenge.

The two programs in which the desired outcomes were modified tended to have very global long-range objectives, such as reducing death from breast cancer, for which many influences other than the program can affect either the incidence of cancer or its mortality rate. Rather than target their annual performance goals directly on the ultimate goal over which they had little control, the respondents said that they identified activities, such as screening for disease, that were known from previous research to be effective in achieving the long-range goals. They used these activities as the basis for specifying annual goals. Thus, the program focused its annual goals, instead, on expanding the delivery of screening, which it can more directly affect.

Approaches to Developing Performance Measures That Reflect Outcomes, Not Outputs

Getting beyond outputs to develop outcome measures was the challenge most often identified as the most difficult in the developing performance measures stage: 18 respondents, representing 17 programs, cited this problem. This challenge did not seem to be as easily reconciled as the

most serious challenge in identifying goals. Two of these respondents reported that they had yet to develop an approach to solving this problem, and none of the respondents thought they had very successfully addressed the challenge. Only 17 percent believed they were mostly successful, whereas most (about 80 percent) believed their approach was somewhat to moderately successful. Respondents finding this challenge particularly difficult had less access to prior studies and assistance from technical staff than the total sample. Two-thirds of these respondents had access to prior studies and technical staff for their approach. All those with access to technical staff reported that they were involved in developing measures that reflected outcomes. (See app. III.)

We found a diverse set of approaches for this challenge; some were focused on conceptual issues, others on measurement issues. (Their approaches and those for the second most often identified challenge in this stage are summarized in table 9.) Several respondents described engaging in conceptual exercises to model the relationships between the program's activities, actors, and objectives to isolate and identify the uniquely federal role. For example, respondents for three programs emphasized the need to recognize the interaction of the federal program and of state and local government efforts. The manager of one of these programs observed that it is difficult for individual agencies at any level of government to specify outcome measures attributable solely to their program because of the interplay among programs at different levels in carrying out program objectives. He thought a more comprehensive measurement model that encompasses federal as well as state and local government activity was needed to identify separate federal outcome measures. He said that his professional community is grappling with the measurement issues involved, but the model has not been developed yet.

Table 9: Approaches Taken to the Most Difficult Challenges in Developing Performance Measures

Challenge	Number of respondents[a]	Approach to developing performance measures
Getting beyond outputs to develop outcome measures	16	Developed a measurement model that encompasses state and local activity to identify outcome measures for the federal program
		Encouraged program managers to develop projections for different funding scenarios
		Conceptualized the outcomes of daily activities
		Used multiple measures that are interrelated
		Developed measures of customer satisfaction
		Used qualitative measures of outcome
		Planned a customer survey
		Involved stakeholders
Specifying quantifiable performance indicators	8	Identified outcome measures used by similar programs
		Conducted a survey
		Involved stakeholders

[a]Number of respondents who identified the challenge as most difficult and had developed an approach to that challenge.

In a second joint federal-state program, it was said to be difficult to gain consensus on a single national outcome because there were conflicting perspectives in the field on the appropriate intervention strategy, and states were thus allowed to develop very diverse programs. One other program used conceptual models or scenario exercises to help program managers broaden their horizons to identify the probable outcomes of their daily activities, asking program staff to imagine what they might be able to accomplish with different levels of resources.

Approaches to the Need to Rely on Others for Data Collection

Using data collected by others was identified as most difficult by more respondents than any other data collection challenge; 11 respondents, representing 9 programs, did so. All reported having developed an approach to this challenge, and most were satisfied with it. More than half the respondents believed their approach was either mostly or very successful.

Respondents reported few resource problems in addressing this challenge. All the respondents reported that prior studies had been conducted, and almost all (90 percent) said that technical staff were available. Most (73 percent) believed the studies were helpful, and those who did used them to a great extent to identify data collection strategies (86 percent) and verify the data (63 percent). All those who had access to technical staff reported that they were involved.

Most of the approaches to this challenge involved either standard procedures to verify and validate the data submitted to the program by other agencies or a search for alternative data sources, as shown in table 10, together with approaches for the next two most frequently identified challenges. For example, to verify data submitted by other agencies, some respondents reported that they had contacted the agency and asked it to correct the data or had hired a contractor to do so. Another respondent reported that to replace existing outcome data that the program had obtained from others, program representatives entered into roundtable discussions with their customers to identify new variables and undertook a special study to seek new data sources and design a composite index of the outcome variables.

Table 10: Approaches Taken to the Most Difficult Data Collection Challenges

Challenge	Number of respondents[a]	Approach to data collection
Using data collected by others	11	Verified and validated the data
		Researched alternative data sources
		Conducted a special study and redesigned a survey to develop new sources of outcome data
		Involved stakeholders
Obtaining baseline data for comparison	9	Created new data elements
		Used data from other agencies
		Developed a customer survey
		Developed an activity-based cost system
		Involved stakeholders
		Provided training
Ascertaining the accuracy and quality of performance data	9	Used a certified automated data system
		Used data verification procedures
		Acknowledged the data limitations
		Provided training
		Used management experience

[a]Number of respondents who identified the challenge as most difficult and had developed an approach to that challenge.

Approaches to Isolating the Impact of the Program

Separating the impact of the program from the impact of other factors external to the program was identified as most difficult by about half of those who rated challenges in the data analysis and results-reporting stage, and several had not resolved it. Fourteen respondents, representing 11 programs, reported having developed an approach, but 5 respondents, representing 5 programs, had yet to do so. Respondents' assessments of the approaches they had developed were modest—28 percent rated their approach as mostly or very successful in meeting the challenge, whereas 44 percent believed they were moderately successful. (These data are provided in app. III.)

Similar to the group at large, prior studies were available to most of these programs, and most of these respondents (68 percent) believed the studies were helpful, even those who had not yet developed their approach. Although fewer respondents had access to technical staff (74 percent), more than 90 percent of them reported that they were involved in addressing this challenge, including some of those with approaches still to be developed. (See app. III.)

Program officials described using a variety of techniques employed in formal evaluations of program impact as well as other approaches to address this challenge, as summarized in table 11. Notably, these techniques were often employed at the subnational level, where the influence of other variables was either reduced or easier to observe and control for. For example, because one such program is well aware that the economy has a strong effect on a loan program's performance, it monitors changes in the economy very closely, but at the regional level. Disaggregating the data to follow one regional economy at a time allows program staff to determine whether an increase in loan defaults in a given region reflects a faltering economy or indicates some problem in the program that needs follow-up. Another program, faced with similar complexities, was said to sponsor special studies to identify its impact at the local level, where it can control for more factors. Since this approach would be too expensive to implement for the entire nation, the program conducts this type of analysis only in selected localities.

Table 11: Approaches Taken to the Most Difficult Analysis Challenge

Challenge	Number of respondents[a]	Approach to analysis
Separating the impact of the program from the impact of other factors external to the program	14	Specified as outcomes only the variables that the program can affect
		Advised field offices to use control groups
		Used customer satisfaction measures
		Monitored the economy at the regional level
		Expanded data collection to include potential outcome variables
		Analyzed time-series data
		Analyzed local-level effects that are more clearly understood
		Involved stakeholders

[a]Number of respondents who identified the challenge as most difficult and had developed an approach to that challenge.

Other programs minimized the influence of external factors on their programs' outcomes through their selection of performance measures. Some programs selected performance measures that are quite proximate to program outputs, permitting a more direct causal link to be drawn between program activities and results. Another program did not have the information it needed to analyze its impacts and settled for measures of customer satisfaction.

Early Implementation Was Assisted by Evaluation Resources

As examples of their agencies' cutting-edge efforts in performance measurement, these programs appeared to have an unusual degree of program evaluation support from within their agencies, as shown in table 12. Despite a 1994 survey that found a continuing decline in evaluation capacity in the federal government, 58 percent of our respondents said they had access to prior evaluations of their program, and 69 percent had access to other studies of their program; 83 percent reported having access to program evaluators or other technically trained staff.[5] Of those with access to program evaluators, 89 percent reported that program evaluators in some way assisted their efforts. Several of the official GPRA

[5]Michael J. Wargo, "The Impact of Federal Government Reinvention on Federal Evaluation Activity," Evaluation Practice, 16:3 (1995), pp. 227-37. An earlier, similar assessment can be found in Program Evaluation Issues (Washington, D.C.: U.S. General Accounting Office, 1992).

pilots were actually run by program evaluation and planning offices. Almost all respondents (96 percent) from large programs (those with annual budgets over $1 billion) reported having access to evaluators, and even 67 percent of respondents from small programs (with budgets under $100 million) reported such access. However, among those with access to evaluators, small programs were less likely than their large counterparts to actually obtain assistance from evaluators (78 percent compared with 95 percent).

Table 12: Respondents' Reported Access to and Use of Evaluation Resources

Evaluation resource	Total sample (percent)	No. of valid cases
Prior studies available		
Program evaluations	58	67
Other studies	69	65
Either	81	67
Prior studies were helpful in		
Defining and setting goals	77	53
Developing measures or planning data collection	81	53
Analyzing data and reporting results	65	48
Evaluation staff		
Available	83	64
Involved	89	56
Evaluation or technical staff were involved in		
Defining and setting goals	80	60
Developing measures or planning data collection	88	60
Analyzing data and reporting results	68	57

Respondents considered prior studies of their program as more helpful in the stages of identifying goals, developing measures, and collecting data (77 and 81 percent) than in the analysis and reporting stage (65 percent). Prior studies were considered most helpful with the tasks of defining program goals, describing the program environment, and developing quantifiable or readily measurable indicators, but least helpful with setting performance targets and explaining program results. Similarly, evaluators and other technically trained staff were said to be most involved in developing performance measures and data collection strategies (88 percent among those with access), particularly in the task of developing quantifiable, readily measurable performance measures, and least involved in the analysis and reporting stage (68 percent).

To develop quantifiable performance measures, for example, one program used a data collection instrument developed in a prior study to collect data on the outcomes of the program on the overall family environment of its target population. An evaluator serving as a consultant to the program identified the data collection instrument. An administrator of another program, who was trained in evaluation methods, used his expertise to develop quantifiable measures for the outcome of a program subject to so many external social and environmental factors that a single performance measure was difficult to isolate. He developed a series of measures that are linked to one another and looked at the overall direction of the measures as the performance indicator. This approach, he suggested, recognized that measuring overall performance is a more complex problem for some programs than looking at a single number or group of numbers.

Yet, it was in the tasks involved in developing performance measures and data collection strategies that respondents were most likely to report they could have used more help: creating quantifiable, measurable performance indicators (56 percent) and developing or implementing data collection and verification plans (48 and 49 percent). When asked why they were not able to get the help they needed, some mentioned lack of time, unavailability of staff, or lack of performance measurement expertise, but more commonly they reported that it was hard to know in advance that evaluators' expertise would be needed (42 percent).

Others were aware that additional research is needed but faced complex measurement issues that staff could not resolve. For example, the respondent whose program is collecting data on family environment outcomes (previously mentioned) needed more dimensions than those provided by the data collection instrument the program was using. The program is conducting exploratory work to identify some of those dimensions. In addition, it still has to determine how to measure the program's long-term effects on parents and children. Another program is looking for sound evidence that services provided to its clients may prevent those families from applying for and receiving more expensive benefits from other public programs. The respondent reported plans to conduct research on this issue.

Conclusions

Seeking to improve government performance and public confidence in government, GPRA established a requirement for executive branch agencies to identify agency and program goals and report on program results. In

reviewing the progress and challenges of selected programs' efforts to complete the analytic steps involved, we found that although agencies have been experimenting with performance measurement for 3 years or more, most have not completed all the tasks required by the Act, and many others are still grappling with the analytic and technical challenges involved. Thus, we expect agencies' full implementation to be an evolving process requiring several iterations to achieve valid, reliable, and useful performance reporting systems. However, we also expect both the agencies and the Congress to benefit from performance measurement as reporting systems are strengthened.

The programs we reviewed are not only volunteers but also have more than average experience with and access to analytical resources in addressing the challenges of performance measurement. Although access to analytic expertise did not solve all these programs' challenges, most of our respondents considered it helpful, and many said they could have used even more such assistance. Thus, with full implementation across the government, more typical federal programs are likely to find performance measurement an even greater challenge, particularly if they do not have access to program evaluation or other analytic resources.

A recurring source of the programs' difficulty both in selecting appropriate outcome measures and in analyzing their results stemmed from two features common to many federal programs: the interplay of federal, state, and local government activities and objectives and the aim to influence complex systems or phenomena whose outcomes are largely outside government control. In such cases, it may be important to supplement performance measurement data with impact evaluation studies to provide an accurate picture of program effectiveness. In addition, systematic evaluation of how a program was implemented can provide important information about why a program did or did not succeed and suggest ways to improve it.

Agency Comments

We discussed a draft of this report with a senior official at OMB. He suggested some technical changes, which we have incorporated.

We are sending copies of this report to the Chairmen and Ranking Minority Members of the Senate and House Committees on the Budget, the Senate and House Committees on Appropriations, and the Subcommittee on Government Management, Information, and Technology, House

Committee on Government Reform and Oversight; the Director of OMB; and other interested parties. We will also make copies available to others on request.

If you have any questions concerning this report or need additional information, please call William J. Scanlon on (202) 512-4561 or Stephanie Shipman, Assistant Director, on (202) 512-4041. Other major contributors to this report are listed in appendix IV.

William J. Scanlon
Director, Advanced Studies and Evaluation Methods

L. Nye Stevens
Director, Federal Management and Workforce Issues

Contents

Figure

Abbreviations

GPRA	Government Performance and Results Act of 1993
OMB	Office of Management and Budget

Objectives, Scope, and Methodology

In order to provide information that may assist federal agencies in meeting the analytic challenges of performance measurement and to help the Congress in interpreting the program performance information provided, we focused our review of agencies' early experiences with performance measurement on three questions:

1. What analytic and technical challenges are agencies experiencing as they try to measure program performance?

2. What approaches have they taken to address these challenges?

3. How have agencies made use of program evaluations or evaluation expertise in implementing performance measurement?

To capture the broad range of performance measurement challenges that federal programs are likely to encounter, rather than to precisely estimate the frequency of those challenges among early implementers, we selected a nonrandom, purposive sample of federal programs that had begun measuring their performance. We based the sample on several factors that we thought might affect their experience. Generally, we selected two programs each from the 14 cabinet departments and from 6 independent agencies—one program that had been designated as an official Government Performance and Results Act of 1993 (GPRA) pilot and another that had begun performance measurement activities on its own or in response to the Office of Management and Budget's (OMB) fiscal year 1998 budget request. Because some agencies had no official GPRA pilot program, 17 of our programs were GPRA pilots, while 23 were not. (See the list of programs we reviewed at the end of this app.) For each program, we attempted to interview both the program official responsible for performance measures and a program evaluator or other analyst who had assisted in this effort. Since no evaluator was identified in some programs, while in others the evaluator was the person responsible for the performance measurement effort, we conducted 68 interviews with officials from 40 programs.

To learn what kinds of technical and analytic challenges agencies were experiencing, we asked these program officials to rate (on a five-point scale) the level of difficulty they had experienced with potential challenges at each stage of the process of developing performance information: identifying goals, selecting measures, collecting data, and analyzing data and reporting results. We identified seven to nine potential challenges for each stage from the literature on performance measurement

and program evaluation and from pretest interviews. We then asked program officials to identify their most difficult challenge in each stage, to describe what approach they took to address it, and to rate (on a five-point scale) how successfully that approach met the challenge. Finally, we asked whether prior evaluation studies and program evaluators (or other technically trained staff), if available, were involved in the various tasks of developing performance information.

Characteristics of the Sample

We selected programs to represent diversity on characteristics that we hypothesized might affect their experience in measuring program performance: program purpose; program funding size; locus of program control at the federal, state, or other level; and program funding through annual or multiyear appropriations. Since the nature of what a program intends to achieve is the basis for any measurement of its results, our first criterion was the program's purpose. To capture the range of activities in the federal budget, we considered three broad program purposes: (1) administering regulations; (2) providing services, including military defense; and (3) developing information, including research and development, and statistical and demonstration programs. Because the smaller programs may have fewer resources to spend on oversight but may also have more clearly focused goals than larger programs, we selected programs with a range of budget sizes.

Additionally, the federal government's level of control over results may often depend on whether it has decision-making authority for program structure, objectives, and type of delivery mechanism. Therefore, we selected a mix of programs whose primary actor is a federal, state, or local agency or some other organization. We also thought budgetary independence might affect how programs responded to the Act's requirements; programs not dependent on the Congress for annual funding might not be as far along.

Finally, we also considered how relevant a program was to the agency's core mission. In some agencies, administrative activities resembling fairly simple processes, such as property procurement and management, were selected as pilots. Because questions about the Act's implementation are concerned with how to measure government's more complex activities, we believed that activities more central to the agency's mission would provide more information about the future of the Act's implementation.

Our sample of pilots was generally similar to the entire population of GPRA pilots in the range of program purposes, but it had a larger proportion of pilots whose locus of control was at the federal level (67 percent) than did the population of all pilots (50 percent). It also had a smaller proportion of pilots with funding under $100 million a year (38 percent compared to 43 percent) (see table I.1). However, our total sample, including pilots and other programs, had the same proportion of federally controlled programs as did the population of pilots (50 percent). It also had somewhat more information-development programs (29 percent compared to 19 percent), fewer regulatory programs (13 percent versus 23 percent), and more large programs with funding over $1 billion (36 versus 24 percent) than the population of all pilots. Most programs are funded by annual appropriations and thus were also the largest share, 82 percent, of our sample. The other programs in our sample either received appropriations for multiple years or were funded for the most part through the collection of offsetting fees.

Table I.1: Characteristics of Our Sample and All Official GPRA Pilot Programs

	GAO sample programs			
Program characteristic	**Pilots**	**Other programs**	**Total**	**Official GPRA pilots**
Program purpose				
Provide services or military defense	57%	58%	57%	59%
Develop information	27	32	29	19
Administer regulations	17	11	13	23
Locus of program control				
Federal	67	37	50	50
State	23	42	34	36
Other	10	21	16	14
Annual budget				
Less than $100 million	38	6	21	43
Between $100 million and $1 billion	31	55	44	28
Greater than $1 billion	31	39	36	24
Appropriations				
Annual	79	84	82	[a]
Multiyear	21	16	18	[a]

[a]Not available.

We found neither an enumeration of agency efforts to measure program performance aside from the official pilots nor a characterization of all federal programs on these dimensions, so we do not know how representative our sample is of the full population of federal programs. However, we believe our sample captures the breadth of federal programs across a range of agencies, purposes, actors, sizes, and types of budget authority.

Data Collection and Analysis

Our survey sought both to characterize the range of analytic challenges that federal programs are wrestling with governmentwide and to obtain descriptions of what they are doing to address specific challenges. To satisfy both objectives, we asked all respondents to do two things. First, we asked them to rate the difficulty of the full set of challenges we hypothesized for each of the four performance measurement stages. This provided us with quantitative data for the portion of the sample that had at least begun each stage. Second, we asked them to nominate one challenge in each stage as the most difficult and to describe, in their own words, why it was difficult and what approach their program had developed to address it. This provided us with qualitative data for each challenge that at least one respondent for a program identified as the most difficult in that stage.

To identify the challenges that our entire sample considered the most problematic, we analyzed all respondents' ratings for each challenge across the four performance measurement stages. To explore why these challenges were problematic, we analyzed the qualitative data available from those who had identified them as their most difficult (in that stage). We then performed a more detailed content analysis of the approach data, for the single challenge in each stage that the largest percentage of respondents nominated as their most difficult. This allowed us to characterize the range of approaches being developed by subgroups responding to the same challenge. Because some respondents from the same program identified different challenges as their most difficult, we reported the results on the basis of respondents rather than programs.

We conducted our work between May 1996 and March 1997 in accordance with generally accepted government auditing standards. However, we did not independently verify the information reported by our respondents.

Table I.2 lists the programs, by agency, included in our review.

Table I.2: Programs Included in Our Review

Agency	Program or function
Agency for International Development	Democracy program area, civil society objective; Population and Health, unintended pregnancies objective
Department of Agriculture	Cooperative State Research, Education, and Extension Service; National Agricultural Statistics Service
Department of Commerce	Information Dissemination: Patent and Trademark Office; National Institute of Standards and Technology laboratories
Department of Defense	Air Force Air Combat Command; Navy Atlantic Fleet
Department of Education	Vocational Rehabilitation State Grant Program; Even Start
Department of Energy	Office of Energy Efficiency and Renewable Energy; science and technology priority area in the Department's performance agreement with the President
Department of Health and Human Services	Office of Child Support Enforcement; Performance Partnerships in Health, Mental Health; Performance Partnerships in Health, Chronic Disease
Department of Housing and Urban Development	Office of the Chief Financial Officer, Departmentwide Debt Collection; affordable housing for low-income renters priority area in the Department's performance agreement with the President
Department of the Interior	U.S. Geological Survey, National Water Quality Assessment Program; Office of Surface Mining Reclamation and Enforcement
Department of Justice	Organized Crime Drug Enforcement Task Force; U.S. Marshals Service
Department of Labor	Occupational Safety and Health Administration; Employment and Training Administration
Department of State	Bureau of Diplomatic Security; International Narcotics Program and Law Enforcement Affairs
Department of Transportation	Federal Highway Administration, Federal Lands Highway Organization; Federal Highway Administration, Federal Aid Highway program
Department of the Treasury	U.S. Customs Service, Office of Enforcement; U.S. Secret Service
Department of Veterans Affairs	Veterans Benefits Administration, Loan Guaranty Service; Veterans Health Administration, medical care programs
Environmental Protection Agency	Acid Rain Program; Air and Radiation Program
Federal Emergency Management Administration	Mitigation budget activity area; National Flood Insurance Program
National Aeronautics and Space Administration	Aeronautics; Human Exploration
National Science Foundation	Science and Technology Centers; Research Projects
Social Security Administration	Entire agency

Overview of GPRA Requirements

The 1993 GPRA, or Results Act, legislation is the primary legislative framework through which agencies will be required to set goals, measure performance, and report on the degree to which goals were met. It requires each federal agency to develop, no later than by the end of fiscal year 1997, strategic plans that cover a period of at least 5 years and include the agency's mission statement; identify the agency's long-term strategic goals; and describe how the agency intends to achieve those goals through its activities and through its human, capital, information, and other resources. Agencies are to identify critical external factors that have the potential to affect the achievement of strategic goals and objectives, include a description of any program evaluations used to establish goals, and set out a schedule for periodic future evaluations. Under the Act, agency strategic plans are the starting point for agencies to set annual goals for programs and to measure the performance of the programs in achieving those goals.

Also, the Act requires each agency to submit to OMB, beginning for fiscal year 1999, an annual performance plan. The first annual performance plans are to be submitted in the fall of 1997. The annual performance plan is to provide the direct linkage between the strategic goals outlined in the agency's strategic plan and what manager and employees do day to day. In essence, this plan is to contain the annual performance goals the agency will use to gauge its progress toward accomplishing its strategic goals and to identify the performance measures the agency will employ to assess its progress. Also, OMB will use individual agencies' performance plans to develop an overall federal government performance plan that OMB is to submit annually to the Congress with the president's budget, beginning with the budget for fiscal year 1999.

The Act requires that each agency submit to the president and to the appropriate authorization and appropriations committees of the Congress an annual report on program performance for the previous fiscal year (copies are to be provided to other congressional committees and to the public upon request). The first of these reports, on program performance for fiscal year 1999, is due by March 31, 2000, and subsequent reports are due by March 31 for the years that follow. However, for fiscal years 2000 and 2001, agencies' reports are to include performance data beginning with fiscal year 1999. For each subsequent year, agencies are to include performance data for the year covered by the report and 3 prior years.

In each report, each agency is to review and discuss its performance compared with the performance goals it established in its annual

performance plan. When a goal has not been met, the agency's report is to explain the reasons why the goal was not met; plans and schedules for meeting the goal; and, if the goal was impractical or not feasible, the reasons for that and the actions recommended. Actions needed to accomplish a goal could include legislative, regulatory, or other actions; when an agency finds a goal to be impractical or infeasible, the report is to contain a discussion of whether the goal ought to be modified.

In addition to evaluating the progress made toward achieving annual goals established in the performance plan for the fiscal year covered by the report, an agency's program performance report is to evaluate the agency's performance plan for the fiscal year in which the performance report was submitted (for example, in their fiscal year 1999 performance reports, due by March 31, 2000, agencies are required to evaluate their performance plans for fiscal year 2000 on the basis of their reported performance in fiscal year 1999). Finally, the report is to include the summary findings of program evaluations completed during the fiscal year covered by the report.

The Congress recognized that in some cases, not all the performance data will be available in time for the March 31 reporting date. In such cases, agencies are to provide whatever data are available, with a notation as to their incomplete status. Subsequent annual reports are to include the complete data as part of the trend information.

In crafting GPRA, the Congress also recognized that managerial accountability for results is linked to managers having sufficient flexibility, discretion, and authority to accomplish desired results. The Act authorizes agencies to apply for managerial flexibility waivers in their annual performance plans beginning with fiscal year 1999. The authority of agencies to request waivers of administrative procedural requirements and controls is intended to provide federal managers with more flexibility to structure agency systems to better support program goals. The nonstatutory requirements that OMB can waive under the Act generally involve the allocation and use of resources, such as restrictions on shifting funds among items within a budget account. Agencies must report in their annual performance reports on the use and effectiveness of any managerial flexibility waivers that they receive.

The Act calls for phased implementation so that selected pilot projects in the agencies can develop experience from implementing the Act's requirements in fiscal years 1994 through 1996 before implementation is

required for all agencies. About 70 federal organizations participated in this performance planning and reporting pilot phase. OMB was required to select at least five agencies from among the initial pilot agencies to pilot managerial accountability and flexibility for fiscal years 1995 and 1996; however, OMB did not do so.[6]

Finally, the Act requires OMB to select at least five agencies, at least three of which have had experience developing performance plans during the initial GPRA pilot phase, to test performance budgeting for fiscal years 1998 and 1999. Performance budgets to be prepared by pilot projects for performance budgeting are intended to provide the Congress with information on the direct relationship between proposed program spending and expected program results and the anticipated effects of varying spending levels on results. To allow the agencies more time for learning, OMB is planning to delay this phase for 1 year.

[6]For information on the managerial accountability and flexibility waiver process, see GPRA: Managerial Accountability and Flexibility Pilots Did Not Work as Intended (GAO/GGD-97-36, Apr. 10, 1997).

Access to and Use of Evaluation Resources

Item	Most difficult challenge in each stage: Translating long-term goals into annual performance goals	Getting beyond outputs to develop performance measures	Using data collected by others	Separating the impact of the program from the impact of other external factors to the program
Number of respondents who selected this challenge as their most difficult	12	18	12	23
Number of respondents who had developed an approach to their most difficult challenge	12	16	11[a]	14[b]
Number of respondents whose approach was still to be developed	0	2	0	5
Number of respondents who had access to prior studies	9	12	11	19
Percentage who considered prior studies helpful	100%	75%	73%	68%
Number of respondents who had access to technical staff	10	12	10	17
Percentage who were assisted by those technical staff	90%	100	100%	94%
Respondents' view of success (percent)[c]				
Minimally successful	0	6	9	17
Somewhat successful	0	28	18	11
Moderately successful	50	50	18	44
Mostly successful	33	17	46	22
Very successful	17	0	9	6

[a]The answer given by one respondent did not match the question format.

[b]Answers given by four respondents did not match the question format.

[c]Percentages may add to more than 100 because of rounding.

Major Contributors to This Report

The following team members made important contributions to this report: Daniel G. Rodriguez and Sara E. Edmondson, Senior Social Science Analysts, co-directed the survey and analysis of agencies' experiences. Joseph S. Wholey, Senior Adviser for Evaluation Methodology; Michael J. Curro and J. Christopher Mihm, Assistant Directors; and Victoria M. O'Dea, Senior Evaluator, provided advice throughout the development of the report.

Related GAO Products

GPRA: Managerial Accountability and Flexibility Pilots Did Not Work as Intended (GAO/GGD-97-36, Apr. 10, 1997).

Performance Budgeting: Past Initiatives Offer Insights for GPRA Implementation (GAO/AIMD-97-46, Mar. 27, 1997).

Measuring Performance: Strengths and Limitations of Research Indicators (GAO/RCED-97-91, Mar. 21, 1997).

Child Support Enforcement: Reorienting Management Toward Achieving Better Program Results (GAO/HEHS/GGD-97-14, Oct. 25, 1996).

Executive Guide: Effectively Implementing the Government Performance and Results Act (GAO/GGD-96-118, June 1996).

Managing for Results: Achieving GPRA's Objectives Requires Strong Congressional Role (GAO/GGD-96-79, Mar. 6, 1996).

Block Grants: Issues in Designing Accountability Provisions (GAO/AIMD-95-226, Sept. 1, 1995).

Managing for Results: Status of the Government Performance and Results Act (GAO/T-GGD-95-193, June 27, 1995).

Managing for Results: Critical Actions for Measuring Performance (GAO/T-GGD/AIMD-95-187, June 20, 1995).

Managing for Results: The Department of Justice's Initial Efforts to Implement GPRA (GAO/GGD-95-167FS, June 20, 1995).

Government Reform: Goal-Setting and Performance (GAO/AIMD/GGD-95-130R, Mar. 27, 1995).

Block Grants: Characteristics, Experience, and Lessons Learned (GAO/HEHS-95-74, Feb. 9, 1995).

Program Evaluation: Improving the Flow of Information to the Congress (GAO/PEMD-95-1, Jan. 30, 1995).

Managing for Results: State Experiences Provide Insights for Federal Management Reforms (GAO/GGD-95-22, Dec. 21, 1994).