

# Using Evaluation to Support Performance Management: A Guide for Federal Executives



Kathryn E. Newcomer  
Professor and Chair  
Department of Public Administration  
George Washington University

Mary Ann Scheirer

The PricewaterhouseCoopers Endowment for  
**The Business of Government**

**About The Endowment**

Through grants for Research and Thought Leadership Forums, The PricewaterhouseCoopers Endowment for The Business of Government stimulates research and facilitates discussion on new approaches to improving the effectiveness of government at the federal, state, local, and international levels.

Founded in 1998 by PricewaterhouseCoopers, The Endowment is one of the ways that PricewaterhouseCoopers seeks to advance knowledge on how to improve public sector effectiveness. The PricewaterhouseCoopers Endowment focuses on the future of the operation and management of the public sector.

**Using Evaluation to Support  
Performance Management:**  
A Guide for Federal Executives

**Kathryn E. Newcomer**  
Professor and Chair  
Department of Public Administration  
George Washington University

**Mary Ann Scheirer**

January 2001

# TABLE OF CONTENTS

<b>Foreword</b> .....	3
<b>Executive Summary</b> .....	4
<b>Introduction</b> .....	7
The Role of the Federal Executive in Performance Management .....	7
Methodology .....	9
<b>Program Evaluation Tools to Support Performance Management</b> .....	11
Requirements for Performance Management .....	11
Types and Methods of Program Evaluation .....	12
Evaluation Uses in the 23 Federal Departments and Component Agencies .....	19
<b>Leveraging Evaluation Capacity to Contribute to Performance Management</b> .....	21
Finding Evaluation Capacity .....	22
Building Evaluation Skills and Knowledge .....	23
Asking the Right Evaluative Questions about Performance .....	23
<b>Organizational Opportunities for Integrating Program Evaluation and Performance Management</b> .....	25
Secure Support from Top Administrators .....	25
Institutionalize Continuous Performance Improvement Strategies .....	26
Use a Variety of Coordinating Processes .....	27
Clarify the Roles of Key Players .....	28
Work Proactively with Congress .....	29
Build Upon Information Technology to Support Performance Management .....	30
Take Advantage of Initial Groundwork for Institutionalizing Performance Measurement .....	30
<b>Conclusions and Recommendations</b> .....	32
<b>References</b> .....	36
<b>About the Authors</b> .....	38
<b>Key Contact Information</b> .....	40

# Foreword

January 2001

On behalf of The PricewaterhouseCoopers Endowment for The Business of Government, we are pleased to present this report by Kathryn E. Newcomer and Mary Ann Scheirer, "Using Evaluation to Support Performance Management: A Guide for Federal Executives."

This report is a timely assessment of how program evaluation is being used in the federal government to support performance management. Newcomer and Scheirer also examine the potential uses of program evaluation to assist federal managers in meeting the reporting requirements set forth by the Government Performance and Results Act.

The report has been prepared as a guide for federal executives responsible for managing and leading government programs. Newcomer and Scheirer describe the many ways in which managers can use program evaluation: support for strategic and program planning, support for improving program delivery, support for accountability, and support for attributing results to agency programs.

In the years ahead, it is likely that pressure will increase for greater accountability from executives on how well their programs are performing and how they are serving the American public. Federal executives will also continue to face the challenge of providing better performance information and data in their strategic and program plans. This report has been prepared to assist them in meeting these challenges. We trust that federal executives will find it helpful as a guide to more effectively using program evaluation as a management tool.

Paul Lawrence  
Partner, PricewaterhouseCoopers  
Co-Chair, Endowment Advisory Board  
paul.lawrence@us.pwcglobal.com

Ian Littman  
Partner, PricewaterhouseCoopers  
Co-Chair, Endowment Advisory Board  
ian.littman@us.pwcglobal.com

# Executive Summary

Federal agencies have now completed their first cycles of planning, measuring, and reporting on programmatic performance required by the Government Performance and Results Act (GPRA) by submitting their first performance reports in the spring of 2000. The analytic work needed to report performance data, to draw apt comparisons, and to probe the validity of the measures raises questions about potential roles for program evaluation to support continued performance improvement. How can evaluators' expertise contribute to this work? Efforts undertaken to report on performance raise many issues concerning roles for evaluation that need continued involvement from agency executives and managers.

This guide draws on a study of the current capacities for and uses of program evaluation to support performance management within 23 large federal departments and agencies (13 departments and 10 component agencies). Interviews with 61 federal managers are summarized to offer advice to federal executives on:

- Program evaluation tools to support performance management;
- Leveraging evaluation capacity to contribute to performance management; and
- Organizational opportunities for integrating program evaluation and performance management.

Program evaluation is defined fairly broadly in this study as "the application of systematic analytic methods to address questions about program operations and results." This approach uncovered more uses and planned uses for evaluation than previously described; official reporting under GPRA may reveal only the "tip of the iceberg" regarding the tremendous amount of analytical work underway in federal agencies as programmatic performance is monitored and probed.

Many stakeholders are now asking good, hard questions about performance. Federal executives need to marshal evaluation resources to ensure that their agencies provide good data. Federal executives should know:

- What program evaluation is, particularly the wide range of its potential uses;
- The array of possible tools that program evaluation and analytical staff can use to support performance measurement and management;
- The level and availability of evaluation capacity in their agencies;
- The evaluation knowledge and skills needed by line managers in their agencies as well as by analytical staff;
- Effective strategies for leveraging evaluation capacity through cross-fertilization of staff and their skills; and

- Opportunities available for marshaling evaluation to support performance management.

## Steps in Performance Measurement

Performance-based management involves more than simply recording measures of program performance and reporting them upwards to oversight bodies and public stakeholders. Several steps are needed to develop and collect performance measures useful for decision making:

- Program stakeholders must come together to reach agreement on strategic and performance objectives and the strategies for achieving them;
- Indicators must be defined for program components that capture program outputs (e.g., the extent of services provided) and/or outcomes (e.g., behaviors of beneficiaries influenced by the services);
- Data sources must be developed or discovered for those indicators;
- Data must be collected with systematic methods, often in multiple jurisdictions (e.g., states, grant-funded projects, etc.);
- Data must be aggregated and reported in user-friendly formats;
- The data must be used by program managers and decision makers to assess and improve results; and
- Data quality must be addressed at every step of its journey from original collection to final reporting.

All of these steps are taken in relation to an agency's program strategies and activities to address the problem or content domain targeted in the agency's mission.

## Evaluation Tools

The tools of program evaluation can contribute substantially to the multiple steps needed for institutionalizing useful performance measurement systems. From the examples provided by respondents, we developed a conceptual framework to help categorize potential uses of evaluation and to encourage fuller use of the broad range of evaluation

tools. The framework suggests that evaluation can be used in all four stages of agencies' work toward performance-based management:

- *Support for strategic and program planning* — for example, for needs assessment, summarizing prior evaluations about “what works,” and to show expected links between strategies and outcomes via logic models;
- *Support for improving program delivery* — for example, by detailing the logic of intended delivery, by using monitoring systems to inform decision making about services, and by using intensive process evaluation to uncover “how” and “why” programs work;
- *Support for accountability* — for example, for developing, reporting, and validating the performance data in GPRA Annual Performance Reports; and
- *Support for attributing results to agency's programs* — for example, using research and evaluation to assess the causal effectiveness of the programs and strategies whose “results” are reported in other documents.

## Agency Capacity for Evaluation

Demands for performance data require analytic capacity and resources often labeled as “program evaluation.” Yet nine of the 13 federal departments contacted for this study had downsized their central evaluation capacity since the 1980s or never had a central evaluation office. The number of staff for program evaluation work has declined steadily from the late 1970s into the 1990s. Few interviewees felt comfortable about their agency's capacity for evaluation. Three themes in leveraging evaluation capacity emerged in interviews:

- Evaluation capacity is difficult to identify when these resources are scattered among operational program offices;
- Building evaluation skills and knowledge is necessary in both staff and line offices, but resources for training are minimal; and
- Line managers need enhanced skills to ask the right kinds of evaluative questions about program performance.

## Organizational Strategies

Underlying organizational processes offer both challenges and opportunities to help integrate evaluation and performance management. We offer seven major organizational strategies that should be addressed by federal executives:

- Secure support from top administrators;
- Institutionalize continuous performance improvement strategies;
- Use a variety of coordinating processes to bring skilled staff together;
- Clarify the roles of key players;
- Work proactively with Congress;
- Build upon information technology to support performance management; and
- Take advantage of initial groundwork institutionalizing performance measurement.

## Recommendations

A variety of recommendations are offered to improve the synergistic relationships between program evaluation and performance measurement to foster performance-based management efforts. A key strategy is to leverage resources available to use evaluative tools in creative ways. The bottom line is that strengthening evaluation capacity and use will enhance the likelihood that the performance management framework being institutionalized via GPRA will result in both improved program management and desired results.

### Make Fuller Use of Program Evaluation Tools and Skills

1. Recognize the diversity of ways that evaluation tools can support performance management, and plan for a broad array of evaluative functions.
2. Conduct an inventory of the evaluation skills of the agency staff directly involved in planning performance measures and reporting them.
3. Seek out staff with evaluation skills when implementing many of the steps in performance management.

4. Provide more explicit guidance for program officers who oversee contractors conducting program evaluations.

### Leverage Capacity for Implementing Performance-Based Management

5. Persuade new political appointees to allocate resources for building greater evaluation capacity.
6. Assess and enhance, if necessary, the evaluation skills of the Office of Inspector General (OIG) staff involved in performance data development and auditing.
7. Search out agency resources that might be re-allocated to support using data for performance management.

### Integrate Program Evaluation and Performance Management

8. Institute and support ongoing teams that bring together evaluators from technical offices, program management, OIG staff, and performance planning and reporting staff.
9. Foster basic performance measurement and evaluation skills as important managerial competencies for all line managers.
10. Plan for continuous improvement in evaluation and performance measurement, building on the groundwork laid in the initial rounds of GPRA work.
11. Publicly reward managers who obtain and use performance data in decision making, for existing as well as new programs.
12. Use evaluation findings in appropriate ways to amplify data about results when reporting to Congress on GPRA requirements.



# Introduction

In the spring of 2000, federal managers completed the first cycle of planning, measuring, and reporting on programmatic performance required by the Government Performance and Results Act (GPRA) by submitting their first performance reports. The efforts undertaken to report on performance raise many issues that need continued involvement from agency executives and managers. GPRA requires agencies to routinely collect data on the level and type of program activities (processes), the direct products and services delivered by the program (outputs), and the results of those activities (outcomes) (GAO, 2000, p. 3). For their performance reports, agencies are expected to analyze performance data, draw apt comparisons, and probe the validity of the measures as they explain the operations and outcomes of their programs.

The field of program evaluation provides a variety of strategies and tools that support the ongoing performance measurement required by GPRA. While it does not specifically require agencies to undertake program evaluation studies, GPRA recognizes the complementary nature of program evaluation and performance measurement. The law requires agencies to describe program evaluations used to establish and revise goals, to include a schedule for evaluations in their strategic plans and to summarize the findings of relevant evaluations in their Annual Performance Reports (GAO, 2000, p. 4). In addition, GPRA requires agencies to provide explanations for why performance targets are not met, another clear opportunity for the application of program evaluation methods. This report suggests

strategies and tools for agency executives to use in meeting these requirements.

This guide draws upon a study of the current capacities for and uses of program evaluation to support performance management within 13 large federal departments and 10 agencies. The experiences of federal managers are summarized to offer advice to federal executives on:

- Program evaluation tools to support performance management;
- Leveraging evaluation capacity to contribute to performance management; and
- Organizational opportunities for integrating program evaluation and performance management.

Tips on using and leveraging evaluation to support performance management are offered. Then recommendations are suggested to identify ways that program evaluation capacity might be leveraged to effectively improve performance measurement and management. Findings and recommendations offered here could also be applicable to agencies in many other jurisdictions facing similar demands for sound performance information.

## **The Role of the Federal Executive in Performance Management**

While GPRA is the major driver of performance measurement and management at the federal level of government, there are many other legislative

**Table 1: Drivers of Performance Management in the Federal Government**

<p><b><i>Pressures External to the Federal Government</i></b></p> <ul style="list-style-type: none"> <li>• International success stories (e.g., New Zealand, UK)</li> <li>• State and local government success stories</li> <li>• The Governmental Accounting Standards Board call for “service efforts and accomplishments” reporting</li> <li>• TQM, quality management, and other initiatives focusing attention on customers</li> </ul> <p><b><i>Executive Initiatives</i></b></p> <ul style="list-style-type: none"> <li>• The National Performance Review/ National Partnership for Reinventing Government</li> <li>• Office of Management and Budget requirements for performance information to accompany budget requests (beginning May 5, 1992)</li> <li>• Executive Order 12862 — To survey customers</li> </ul> <p><b><i>Legislative Initiatives</i></b></p> <ul style="list-style-type: none"> <li>• Chief Financial Officers Act of 1990</li> <li>• Government Performance and Results Act of 1993</li> <li>• Government Management and Reform Act of 1994</li> <li>• Chief Information Officers Act of 1996</li> <li>• Federal Financial Management Improvement Act of 1996</li> <li>• Many other laws requiring performance measurement in specific policy areas passed after 1995.</li> </ul>
--

requirements to report on performance, for agencies at the federal as well as state and local levels of government. Table 1 lists the many factors pressing federal agencies to focus on performance reporting (Newcomer, 2000). As indicated, the examples of other countries and jurisdictions, as well as a continuous series of executive and legislative actions, have initiated a culture of “performance-based management.” Traditional political and interest group pressures on decision making are being supplemented, but not replaced, by demands for data about the operations and results of government programs.

These demands for performance data require analytic capacity and resources often labeled as “program evaluation.” However, support for program evaluation capacity across federal agencies has actually declined during the last two decades (Newcomer, 1994). The number of full-time staff dedicated to conducting program evaluation work within federal agencies declined fairly steadily from the late 1970s into the 1990s. Nine of the 13 federal departments contacted for this study downsized their central evaluation capacity since the 1980s or never had a central evaluation office. Recognizing the diminished ability of federal agencies to assess systematically how well programs are working, the authors of GPRA state they inserted language about the use of program evaluation in GPRA to stimulate interest and resources for evaluation work across agencies (NAPA, 2000).

Given the generally low capacity for program evaluation in most federal agencies, and the high demands for valid and reliable evidence of the results of federal efforts required by GPRA, how might federal executives rise to the analytical challenges thrust upon them? We conducted this study to identify ways that federal managers are using program evaluation to contribute to these analytical challenges. The objectives of this work were the following:

- To document ways that program evaluation strategies and tools are being employed to support implementation of GPRA, and
- To offer recommendations for improving the linkages between program evaluation and performance management in the federal government.

The current environment in the federal government requires federal executives to measure and report on programmatic performance. Performance-oriented political appointees, staff of the Office of Management and Budget (OMB), congressional committees, and interest groups are asking good, hard questions about performance. Federal executives need to know how to marshal resources to ensure that their agencies provide good data. Federal executives should know:

- What program evaluation is, particularly the wide range of its potential uses;
- The array of possible tools that program evaluation and analytical staff can use to support performance measurement and management;
- The level and availability of evaluation capacity in their agencies;
- The evaluation knowledge and skills needed by line managers in their agencies as well as by analytical staff;
- Effective strategies for leveraging evaluation capacity through cross-fertilization of staff and their skills; and
- Opportunities available for marshaling evaluation to support performance management.

Program evaluation is defined fairly broadly in this study as the application of systematic analytic methods to address questions about program operations and results. In fact, the federal officials interviewed were encouraged to explain how they view program evaluation. This approach uncovered more uses and planned uses than have been reported in published documents such as the FY 1999 performance reports. Our interviews suggest that official reporting under GPRA may reveal only the “tip of the iceberg” regarding the tremendous amount of analytical work underway in federal agencies as programmatic performance is monitored and probed.

**Methodology**

This report draws upon exploratory research in 13 major departments of the federal government, plus 10 of their component agencies. Federal agency evaluators, and staff members in offices of Inspector General (OIG), policy and planning offices, and

**Table 2: Federal Agencies in the Study (with abbreviations)**

<p>Department of Agriculture (USDA) Food and Nutrition Service (FNS)</p> <p>Department of Education (ED) Office of Student Financial Assistance Programs (SFA)</p> <p>Environmental Protection Agency (EPA)</p> <p>Department of Health and Human Services (HHS) Health Resources and Services Administration (HRSA) Centers for Disease Control and Prevention (CDC)</p> <p>Department of Housing and Urban Development (HUD)</p> <p>Department of Interior (DOI)</p> <p>Department of Justice (DOJ) Bureau of Prisons (BOP) National Institute of Justice (NIJ)</p> <p>Department of Labor (DOL) Employment and Training Administration (ETA)</p> <p>National Aeronautics and Space Administration (NASA)</p> <p>Social Security Administration (SSA)</p> <p>Department of Transportation (DOT) National Highway Traffic Safety Administration (NHTSA) Maritime Administration (MARAD) Research and Special Programs Administration (RSPA)</p> <p>United States Agency for International Development (USAID)</p> <p>Department of Veterans Affairs (VA)</p>
---

budget offices involved with implementing GPRA were interviewed between February and June of 2000. Letters were sent to each interviewee in advance explaining the project, assuring confidentiality, and requesting participation. The interview protocol was often faxed to each interviewee after appointments for in-person or telephone interviews were confirmed. The interview protocol contained mostly open-ended questions developed from the primary research questions for the study. Pre-test interviews to improve the instrument were undertaken with highly experienced evaluation managers within three agencies.

Agencies and respondents were selected using a snowball sampling technique. Initial contacts were made with agencies with track records of experience in evaluation. Other federal agencies were included to broaden the range of federal programs covered. There should be no inference drawn from this study that agencies excluded are not using evaluation to support implementation of GPRA. The agencies included in the study are listed in Table 2. For many agencies, staff in one or more operating units within the agency — e.g, the National Highway Traffic Safety Administration and the Centers for Disease Control — were also interviewed. The operating units were selected based upon the recommendations of central departmental staff, and were typically recommended because of the noteworthy evaluation work they have undertaken.

Thirty-seven interviews were conducted in 23 federal departments and agencies; the interviews involved a total of 61 federal officials, as some included more than one person. The researchers sought out those who had been actually involved in preparing strategic plans, performance plans, and performance reports required under GPRA.

# Program Evaluation Tools to Support Performance Management

Prior discussions of performance measurement and program evaluation often contrast these functions as being *different strategies* for collecting and using information about program performance. For example, the General Accounting Office (GAO) has been active in stimulating and monitoring performance management and its relationships with program evaluation. GAO defines performance measurement as “the ongoing monitoring and reporting of program accomplishments, particularly progress towards pre-established goals.” It states that “program evaluations are individual systematic studies conducted periodically or on an ad hoc basis to assess how well a program is working” (USGAO, 1998).

However, our research revealed that many program managers view their performance measures as a major means of programmatic evaluation, and do not make the conceptual distinctions implicit in the GAO definitions. Instead, evaluative functions are pervasive in the developmental work supporting performance measurement, including tasks such as defining the program outcomes to be measured, designing or discovering measuring tools for these outcomes, collecting valid data, analyzing these data, and presenting results in formats useful to a variety of audiences. But some agencies are not viewing these tasks as part of program evaluation,

and often do not assign them to staff members with skills in these functions. By adopting a broader definition of program evaluation, one which includes performance measurement as a relevant application of evaluators’ skills, agencies might integrate more fully their analytic capacities into their work in implementing GPRA.

## Requirements for Performance Management

Performance management involves more than simply recording measures of program performance and reporting them upwards to oversight bodies and public stakeholders. Several steps are needed to develop and collect performance measures that can be useful to management in decision making:

- Programmatic stakeholders must come together to reach agreement on strategic and performance objectives and the strategies for achieving them;
- Indicators must be defined for program components that capture program outputs (e.g., the extent of services provided) and/or outcomes (e.g., behaviors of beneficiaries influenced by the services);
- Data sources must be developed or discovered for those indicators;

- Data must be collected with systematic methods, often in multiple jurisdictions (e.g., states, grant-funded projects, etc.);
- Data must be aggregated and reported in user-friendly formats;
- The data must be used by program managers and decision makers to assess and improve results; and
- Data quality must be addressed at every step of its journey from original collection to final reporting.

All of these steps are taken in relation to an agency's program strategies and activities to address the problem or content domain targeted in the agency's mission.

A central theme of performance measurement is to collect useful evidence, usually quantitative data, about the delivery and results of agency actions. This process is often called evidence-based management in the health services context. The tools of program evaluation and the skills of program evaluators can contribute substantially to the multiple steps needed for institutionalizing useful performance measurement systems. This section discusses the many ways that evaluation strategies, tools, and studies can support performance management, going far beyond a narrow view of evaluation that focuses primarily on impact research.

Federal agency officials suggested that evaluation is not only a set of specific, discrete studies of programs, but is an ongoing function served by both performance measurement and more targeted studies. Interviewees acknowledged that focused evaluation studies are needed to provide more in-depth understanding of program implementation, to supply evidence of the causal links assumed to connect outputs with intended outcomes, as well as to examine the contextual background for ongoing performance data. Performance measurement provides the central core of routinely collected evaluative data about program operations and outcomes. Additional evaluation studies provide complementary evidence and logical frameworks for increasing the depth of knowledge for interpretation of the performance data.

## Types and Methods of Program Evaluation

Over the past 30 years, evaluators have developed an extensive array of approaches and methods to address a variety of client needs and situations. Methods for process evaluation, theory-based evaluation, impact assessment, and participative evaluation, among others, all provide alternative approaches to data collection and analysis. While this array of evaluative approaches is sometimes bewildering to program managers, it does provide a diversity of tools that can be applied to various tasks needed for developing and using performance measurement systems for evidence-based management.

To illustrate the varied ways that program evaluation can contribute to the many processes involved in performance management, we developed the conceptual framework shown in Table 3. The columns describe major stages in the development and use of systematic information for performance management, including:

- **Support for strategic and program planning** — for GPRA, developing Strategic Plans and Annual Performance Plans;
- **Support for improving program delivery** — the heart of program management, by using the planned strategies and data feedback to inform decision making on delivering services, designing and implementing regulations, and providing grants for research or service delivery;
- **Support for accountability** — reporting the performance data in GPRA Annual Performance Reports and other reports to oversight bodies, such as Congress and the public; and
- **Support for attributing results to agency's programs** — providing research and evaluation to analyze the causal effectiveness of the programs and strategies whose "results" are reported in other documents.

The stages of performance management occur in cycles in most agencies, and may overlap (particularly the work involved in attributing results to agency programs, which may be ongoing). For example, the officials we interviewed had just

**Table 3. Framework Showing Potential Uses of Program Evaluation in Support of Performance Management**

<b>Stages in the Use of Performance Information for Performance Management</b>				
<b>Types and Methods of Program Evaluation</b>	<i>Support for Strategic and Program Planning</i>	<i>Support for Improving Program Delivery</i>	<i>Support for Accountability</i>	<i>Support for Attributing Results to Agency's Programs</i>
<i>Conceptual Development</i>	<p><b>a.</b> To show expected links via logic models</p> <p><b>b.</b> For policy analysis of options available</p>	<p><b>a.</b> Use of logic models to show intended delivery</p> <p><b>b.</b> Use of prior literature about program strategies</p>	<p><b>a.</b> To show links among inputs, outputs, and outcomes with logic models</p>	<p><b>a.</b> Use of theory-based evaluation</p> <p><b>b.</b> Use of logic models</p>
<i>Development of Evaluation Methods and Quality Control</i>	<p><b>c.</b> To develop meta-analysis methods for summarizing findings from previous evaluation and research</p>	<p><b>c.</b> To develop data systems for monitoring program delivery and management</p>	<p><b>b.</b> To develop performance measures and data sources</p> <p><b>c.</b> To verify and validate data</p>	<p><b>c.</b> To develop evaluation research designs and measurement tools</p>
<i>Use of Data Systems</i>	<p><b>d.</b> For needs assessment, via population-based data and periodic surveys</p> <p><b>e.</b> Setting targets for performance measures</p>	<p><b>d.</b> Use of MIS data to monitor diverse program sites or states</p> <p><b>e.</b> Geographic-based analysis</p>	<p><b>d.</b> Use of MIS to produce output and outcome data</p>	<p><b>d.</b> By causal modeling (e.g., time-series and regression models) using MIS data</p>
<i>Process and Impact Evaluation Studies</i>	<p><b>f.</b> To summarize findings from prior evaluations and research: "What works?"</p>	<p><b>f.</b> For process and implementation evaluation (one-time or periodic)</p>	<p><b>e.</b> For descriptive outcome data (e.g., using surveys)</p>	<p><b>e.</b> For impact evaluation, with experimental or quasi-experimental design</p> <p><b>f.</b> To analyze plausible associations; pattern matching</p>

## Definitions of Types and Methods for Program Evaluation

*Conceptual Development* — Methods for detailing the pathways by which programs are intended to work.

*Development of Evaluation Methods and Quality Control* — Detailing the procedures and specifications to be used for collecting data, for evaluation study designs, and for verification and validation of data quality.

*Use of Data Systems* — Using data from available statistical indicators (e.g., state-collected “vital statistics” or regularly collected surveys), from program-specific Management Information Systems, and/or from Geographic Information Systems.

*Process and Impact Evaluation Studies* — Systematically conducted assessments, usually on a one-time basis, of the activities or interventions undertaken by a program, and/or of the results attributable to that program.

completed their first Annual Performance Reports and were starting to revise their Strategic Plans. They were explicitly building on their prior work and trying to improve it, using feedback from both internal and external sources. The rows of Table 3 summarize major types of program evaluation that can contribute to each stage in performance management (see accompanying box for definitions). The cells of Table 3 then illustrate the many and varied potential uses for evaluative information in support of performance management. These diverse uses illustrate our broad definition of evaluation as “the application of systematic analytic methods to address questions about program operations and results.”

The following section details the multiple ways that evaluation approaches are being used to support performance management efforts, to amplify the suggestions in Table 3. A “menu of approaches” is offered that agency executives, program managers, and their partner agencies might consider when planning data collection strategies that would provide useful information to support performance

management. As will be evident from the variety of examples described below, these uses of evaluation tools are far more diverse than the traditional distinction between evaluation and performance measurement would suggest.

### Conceptual Development — Using Logic Models and Program Theory

Evaluation tools can support the conceptual work needed to describe the agency’s intended programs and their strategies. Program evaluators (e.g., Wholey, 1979, 1987; McLaughlin and Jordan, 1999) have long advocated the use of logic models to show the flow of program operations, including:

- resources (inputs) and/or initial assumptions;
- implemented program activities;
- program outputs (products); and
- short-term, intermediate, and long-term outcomes.

GPRA legislation and guidance from the General Accounting Office support this focus by suggesting that agency strategic plans and performance measures should include inputs, outputs, and outcomes to show a logical sequence of strategies and results, as well as their measures (U.S. GAO, 1997).

Further, the use of logic models encourages connections to theory-based evaluation (Bickman, 1987; Chen, 1990; Weiss, 1998) by detailing the logic and assumptions connecting the activities delivered by a program and its expected results. Program staff can work with stakeholders to develop logic models and hypothesized program theories to assess the plausibility of expectations for program goals and results. Such conceptual assessment also encompasses the likelihood that intended delivery will be implemented, and that the resources and other requirements will be available to enable those program activities to take place.

Focusing on the conceptual background underlying a program, such as by developing a logic model, can be a valuable tool for agency use in gaining agreement among stakeholders on the intended program, and in clearly communicating program results as a logical sequence among program measures and targeted outcomes. For some measures,



### **Using a Program Flow Model at the Department of Education**

A senior administrator in the Student Financial Aid program at ED developed a 13-step logical flow model of students' decision processes in obtaining higher education and financial aid, including the occurrence of loan repayment problems. ED programs that affect each step were listed. The model occupied a whole wall of the administrator's office, and had been iteratively developed using Post-it notes to add or delete specific elements suggested by program managers or other stakeholders. This administrator referred to the model when discussing program strategies and alternative interventions with program managers, particularly to identify likely intervention points that could make a difference in the outcomes of student financial aid programs.

particularly agency outputs, agencies ought to be able to specify targeted performance rather precisely. For longer-term outcomes or "results," targets are likely to be less precise, because many other partners and environmental factors affect those outcomes, and agencies have less immediate control over them. A logic model helps to show the "chain of results" between an agency's concrete activities and its long-term goals or purposes.

Rather surprisingly, we found only a few explicit examples of the use of logic models in GPRA plans and performance measures. Much of the initial guidance to agency staff on developing performance measures focused heavily on moving beyond outputs to specify and develop measures of outcomes. Perhaps agency staff who developed these documents did not feel it was necessary to show the linkages among their measures or the "theory" underlying their strategies. However, some analytic staff members indicated the need for their agencies to focus on these links in the future, particularly in programs that involve state and local partners.

In areas where there has been a long history of measurement, such as in the National Highway Traffic Safety Administration (NHTSA), there is a higher level

of comfort with complex logic models. Counting highway traffic casualties has been part of work as usual at NHTSA since it was founded, and staff can use complex models with facility. In other instances, the need to re-engineer program processes to focus agency efforts on feasible intervention points has pushed staff to such an analytical approach, as illustrated by the program flow model for the Department of Education's student aid program.

### **Development of Evaluation Methods and Quality Control**

A second contribution that evaluation tools can make to performance measurement is in the development of valid indicators and their associated data collection methods, as well as data quality control methods in ongoing data collection. Development of measures is certainly a traditional role for evaluators, a function recognized by most of those interviewed. Evaluators who specialize in specific content areas can often serve as the "bridge" to outside content area experts to help agency officials select the most up-to-date and valid measures to operationalize an intended performance construct. For example, a Healthy Eating Index used by the Department of Agriculture's Food and Nutrition Service (FNS) was developed over a three-year period, with primary data collection from a nationally representative survey of food intake, along with the involvement of diverse stakeholders such as academics and advocates.

Further, evaluators' technical expertise can often help in specifying the details of data collection methods needed for reliable data. Evaluators and other social researchers have developed professionally approved methods for a wide range of data, such as:

- Collecting data in sample surveys;
- Systematically recording field observations;
- Conducting key informant interviews;
- Entering and maintaining data via handheld computers or local agency management information systems; and
- Collecting qualitative data systematically, with methods such as focus groups, expert panel judgments, and case studies.

In addition, evaluators frequently are experienced in delivering the training needed by agency staff and administrators to ensure the integrity of data collection. These data development and collection methods can be usefully applied to performance measurement when evaluators and agency officials work closely together to integrate needed measures with technically appropriate methodologies.

Prior evaluation and research results can also help agencies to set realistic targets, for example, in GPRA Annual Performance Plans. Prior evidence might be available about the extent of change achieved by use of a specific agency strategy, the time spans needed to achieve those changes, and the extent of variability to expect among delivery sites. Such data can help an agency to be realistic in setting specific targets for its performance measures in each annual plan by linking its strategies and outputs to the prior evidence on the extent of change (“results”) likely to be achieved. Use of evidence can also help in communicating to key stakeholders a realistic view of what is achievable within a short time period in moving toward the long-term desired results. For example, the U.S. Coast Guard has been a leader in analyzing data on marine safety to ensure the managers understand how to interpret trends as they set performance targets, and data-driven management has caught on across the agency.

A further use of program evaluation skills is to assess the quality and integrity of data used for performance measures. For performance data to be credible to users, and to withstand potential criticism, the data must be documentably unbiased and relevant for assessing the agency’s performance. While many agencies have collected substantial archives of data over time, several administrators acknowledged that data quality characteristics were often questionable or unknown. Problems with data collected by diverse partners at multiple levels can take many forms, such as the following:

- Definitions of data elements may be unclear, so that local data collectors interpret them differently and thus enter data differently;
- Training and/or supervision of local data collection may be lax, resulting in unreliable initial data recording or entry;

- Different jurisdictions may have used different time frames when supplying data to a federal office, such as their varying “fiscal years,” which makes it difficult to aggregate the data into national performance indicators; and
- Data processing systems may have used different standards, or be differentially accurate, in the processing and transmission of data sets across jurisdictions or data systems.

Evaluators with appropriate technical expertise can assist in both identifying the nature and implications of such problems and in suggesting the “fixes” or training needed to avoid similar future data problems. For example, the Department of Education is using both evaluation and program staff input, as well as major stakeholder input from states, to redesign the data to be submitted by states concerning their elementary and secondary educational activities in an “Integrated Performance and Benchmarking System.”

#### **Use of Data Systems**

Performance management can be supported in many ways by regularly collected data systems, including publicly available statistics, management information systems (MIS), and Geographic Information Systems (GIS). Even for the initial stages of strategic planning and developing program strategies, existing statistical data sources such as census data or vital statistics can be extremely useful. They can provide data for needs assessment, for estimating the scope of a target population, and for showing the underlying recent trends for potential key outcome indicators. Analytic skills are likely to be needed to identify appropriate existing data sources and to pull out the needed specific information from the broader data archives.

Management information systems are often used to provide periodic data about programmatic activities and outcomes. These systems house evaluative data that are rightly viewed as integral elements of program evaluation, as well as supplying the externally reported performance measures. Monitoring program delivery to describe what is provided to whom has traditionally been a key component of process evaluation (Scheirer, 1994). This approach has recently been extended to evaluation in numer-

### **A Multi-Purpose Management Information System (MIS) at the Department of Labor's Employment and Training Administration (ETA)**

ETA is creating an information system for its new Workforce Incentives Act programs which will also integrate data on participants from its multiple employment assistance programs, including the US Employment Service, Job Corps services and others. This MIS will be used for GPRA and other reporting requirements, will be accessible to its state contributors, and will allow ETA to assess the relationships among program participation, participant characteristics, and outcomes.

ous local service agencies via the United Way's influential *Measuring Program Outcomes* guide and associated technical assistance (United Way of America, 1996). Those advocating outcomes monitoring are extending performance measurement to capture results of program delivery rather than simply counting program actions or such things as persons served (Hatry, 1999).

In some cases, data originally developed for evaluative and monitoring purposes are used by agencies as performance indicators for key outcomes. For example, the Department of Housing and Urban Development (HUD) is using data from the American Housing Survey (AHS), conducted biannually for HUD by the Bureau of the Census, for several GPRA indicators. These include the number of households with "worst case" housing needs and a new index of the number of housing units in "substandard condition." Its analysts drew on several years of analytic work with the AHS to develop a composite measure of "substandard condition."

Other agencies are developing integrated management information systems for performance management that pull together data about program operations, outcomes, and results, often from diverse data sources. Analysts intend to use these MIS resources not only for descriptive data reporting for GPRA, but also for more in-depth data analyses usually considered "social research" or "program

evaluation." These data can be used for multivariate modeling to examine the linkage between variations in program strategies and the outcomes achieved in diverse locations, with controls for other differences in the environment or client characteristics of those locations. When enough data points are available in such systems, data collected over time to show descriptive trends can also be used for more detailed program modeling, such as assessing the impact of a legislative or policy change using an interrupted time-series design.

Another innovative tool is the use of Geographic Information Systems to examine the co-occurrence of programmatic and other indicators (e.g., demographic characteristics, SES statuses, housing characteristics, etc.) by location, such as states, counties, or ZIP codes within a city. For example, the Department of Justice is using geographic information to relate the deployment of federal law enforcement resources among federal judicial districts to caseloads of its component agencies in the same districts, in order to help optimize resource allocation.

#### **Process and Impact Evaluation Studies**

Processes of strategic planning, indicator development, and setting targets all require grounding in the evidence produced by prior program evaluations and applied research in the relevant content areas. For agency strategic plans to be realistic, they need to be based on the cumulative evidence concerning "what works" for achieving the intended outcomes. For example, the Food and Nutrition Service (FNS) conducted an observational study of the nutrition actually supplied in the facilities funded in its Child and Adult Care program, to establish feasible performance measures for this program and to determine whether revised guidelines were needed. The study showed that many care givers provided good nutrition to their children or adult beneficiaries, and FNS made a policy decision that further regulations were not needed.

Evaluation tools for the synthesis of prior studies may be needed for this task, particularly meta-analysis and other methods for reviewing and assessing prior literature. If the literature review is combined with a program logic model, the evaluator can indicate the nature and strength of available evidence underlying causal arrows in the model.

For example, the U.S. Agency for International Development (USAID) used basic research methods to establish the link between the presence of medically trained birth attendants and lower maternal mortality in developing countries. This evidence then justifies AID's use of the indicator "percentage of births attended by medically trained personnel" as a measure of improved maternal health systems, rather than using less reliable data about long-term maternal mortality from diverse countries' government information systems.

Another tool from program evaluation is the use of process evaluation methods to amplify the descriptive information in performance measures, which usually simply describe *what* is occurring within a program, but not *how* or *why*. Evaluation methods for assessing the processes involved in implementing the program in diverse delivery sites, as well as the extent and accuracy of delivery, can help to illuminate how the program operates and, therefore, why performance targets are or are not achieved. This more in-depth information can be used to help improve program results by suggesting needed changes in delivery strategies to increase fidelity to the intended program components. This type of evidence might already be available from prior evaluation studies that provide detailed descriptions of implementation variations. Or, as performance measurement data are obtained, they may raise questions concerning variations among delivery sites that could call for conducting a new process evaluation. For example, international data from USAID's ongoing demographic and health surveys began to show lower rates of childhood immunization in several countries. This reduced implementation prompted the agency to boost its programming in the affected countries to increase immunization rates, a key indicator for its GPRA reporting on children's health status.

Implementation evaluations can use a variety of data collection strategies, including:

- Conducting site visits and case studies of diverse sites;
- Analyzing more detailed data about variations in intervention delivery (perhaps from an ongoing MIS or GIS);

- Holding focus groups to obtain in-depth understanding about program operation, such as from the perspectives of program clients; and
- Interviewing key informants such as local delivery staff or knowledgeable local observers about factors facilitating or impeding local implementation.

The Department of Veterans Affairs (VA) has been a leader in applying these evaluation strategies to clarify understanding of how their educational benefits program is working. They are currently undertaking similar evaluation work in their cardiac care and survivors' benefits programs.

Many interviewees raised concerns about attributing their reported "results" to their often limited program interventions when multiple factors influence these broad societal goals. As Michael Hendricks has aptly noted, attribution is the "Achilles' heel" of performance measurement (Hendricks, 2000). That is, routinely collected data about results allegedly affected by governmental efforts, such as improved student achievement scores, reduced traffic fatalities, and reduced air-traffic "near collisions," are usually not sufficient to "prove" that the governmental actions caused improvements in the conditions monitored. Impact evaluations can be used to address this problem by providing evidence concerning the hypothesized links between agency outputs and program impacts, for example, by using an experimental or quasi-experimental design. Several agencies reported their intentions to expand their use of causal impact studies to support their reporting of results. But such studies are often time-consuming and expensive. Further, the complexity of the statistical methods required for these evaluations often makes it difficult for program managers to assess and use these tools.

Program evaluation professionals have also suggested less rigorous strategies to address the attribution dilemma (Reynolds, 1998). Larry Cooley's ideas about "plausible association" between program efforts and performance help by focusing attention on the choices that must be made about an appropriate standard of evidence in drawing inferences about attribution (Cooley, 1994). John Mayne has suggested some helpful ways to appro-

propriately interpret performance data; for example, he places responsibility on the managers closest to the data to reduce the uncertainty, provide considerable evidence, offer a credible picture of attribution, and acknowledge external (rival) factors in providing a credible performance story (Mayne, 1999). The Department of Transportation (DOT) drew accolades from the Congress and the General Accounting Office for its FY 2000 Performance Report, in part because external factors also affecting monitored behaviors were acknowledged in credible performance stories that the report conveyed in many areas of transportation safety.

Addressing the challenge of causal inference in reporting outcomes of public programs is definitely an area where program evaluation offers many conceptual strategies and tools. To ensure that performance measurement is implemented and reported to support a plausible association between programs and the outcomes measured, targeted evaluation work can be invaluable. For example, evaluations of variability in implementation across sites and longitudinal trends in program participants' behaviors can be used to help make the case for attribution of observed improvements to program actions by tracing the patterns of outcomes in relation to a hypothesized causal sequence (Marquart, 1990). If several waves of performance measures are available, and systematic interventions are undertaken between waves of data, then examining differences among sites before and after the intervention may contribute evidence of the program's contribution to the results achieved (Scheirer, in press, 2000). Evaluations targeted at the linkages identified in program logic modeling efforts can increase confidence that the programs are working.

On the other hand, impact evaluations that systematically trace the implementation and/or the effects of program activities are extremely helpful when performance data indicate that programs are not working. In fact, the Government Performance and Results Act requires agencies to employ program evaluation studies to explain instances where performance fails to achieve targets. While the legislative mandate focuses attention on learning why programs fail, targeted evaluation work can also advance knowledge of when and why programs work.

## Evaluation Uses in the 23 Federal Departments and Component Agencies

As indicated above, we found many examples of a diversity of types of evaluation used to support performance management among the agencies in this study. To provide an indication of the scope of evaluation use, we tabulated the agencies by their types of use. The department or agency was included as "used" or "planned to use" for each type of evaluation if any of those interviewed at that agency mentioned an example of that type of use. The example of use could either be one or more studies cited in one of their GPRA documents or an example that supported their GPRA work but was not explicitly cited in their reports.

Nearly half of the agencies (9 to 11 of the 23) had used evaluation to support the central processes for performance measurement, and 4 to 7 additional agencies planned to use evaluation for these purposes:

- Defining expected performance and targets;
- Developing indicators and data collection methods;
- Measuring outcomes; and
- Assessing program implementation and delivery.

Fewer agencies reported using or planning to use logic models (7 agencies), using or planning to use evaluation to assess the quality of their performance data (10 agencies), or conducting/planning impact evaluations to attribute their results to program interventions (12 agencies). Thus, many agencies were familiar with a variety of types of evaluation use, but these were often just one or two examples in an agency with multiple programs and indicators (sometimes hundreds of them).

The infrequent use of logic models was somewhat surprising to us, as this tool has been widely discussed in connection with GPRA processes. Several respondents who were familiar with logic models but had not used them in their GPRA documents mentioned that logic modeling seems more appropriate for a specific program with a defined set of activities. In contrast, their GPRA plans and indica-

tors are often for broad outcomes that cross-cut more than one of their funded programs. Further, sometimes their managers were reluctant to show outcomes in a logic model that their programs could not directly control, the “attribution for results” problem discussed earlier.

These study findings indicate that many agencies can and do use evaluation when it is available. Nearly all interviewees believed in the value of systematic evidence to support their agency’s work and wanted to apply more evaluation expertise than their agency is currently supporting. Many cited concrete examples of work on specific evaluations or management information systems that would provide expanded information in the future. While nearly all are constrained by their agencies’ limited capacities for evaluation activities, many described a number of management strategies being used to foster the use of program evaluation.

# Leveraging Evaluation Capacity to Contribute to Performance Management

Evaluation capacity was viewed fairly broadly in this study. The components of “capacity” can include the numbers and skills of available evaluation staff, the extent of funding for contracting evaluation studies, and the organizational support for workforce training and professional development to enhance evaluation knowhow. Table 4 provides a broadbrush picture of how respondents in this study assess the current status of evaluation capacity in their agencies. As noted in the table, respondents from central departmental staff offices were asked about staffing at their level, while staff in operating units were asked about analytical staff only in their own units. Respondents were asked about analytical staff available to assist with program evaluation and not to confine their assessment to “evaluation” offices. Only three of the 13 federal departments contacted had centrally located evaluation offices, with three others in the process of developing this capacity. As indicated in Table 4, few agencies were viewed as “comfortable” in any of the dimensions of capacity, and resources to support training were particularly lacking.

The relatively higher level of satisfaction with current evaluation staffing levels expressed by staff in operating units than by their counterparts in central departmental offices reflects, in large part, the purposive agency selection technique used. That is, the operating units were typically selected due to their more

extensive experience with evaluation. For example, the National Highway Traffic Safety Administration was established in 1966 to measure and report data on highway safety, and its staff is comprised largely of statisticians and other professionals quite comfortable with evaluation methodology.

On the other hand, respondents in centralized departmental offices typically reported that the level of both resources and evaluation skills and/or appreciation varied substantially across their departments. In most large federal departments, some operating units may possess both evaluation staff and commitment for evaluation, but typically that is not the case across the entire department. As GPRA reporting documents are usually prepared by a central office, the evaluation work performed by operating units may or may not be reflected in the department-level documents.

Agency staff who are familiar with GPRA requirements for agencies to document their use of program evaluation have typically become aware of their organization’s evaluation capacity. Three themes in leveraging evaluation capacity emerged in interviews. Our respondents were very forthright in noting that:

- Evaluation capacity is difficult to identify;

**Table 4: Current Status of Program Evaluation Capacity in the Agencies Profiled**

	<i>In Need of Improvement</i>	<i>Mixed</i>	<i>Comfortable</i>
<i>Resources</i>			
Adequacy of # of FTEs in Centralized Policy/Evaluation Offices <sup>a</sup>	6	5	2
Adequacy of # of FTEs in Decentralized Policy/Evaluation Offices Contacted <sup>b</sup>	3	3	5
Budget for Contracting-Out Evaluation Work <sup>a</sup>	5	5	3
<i>Workforce and Professional Development</i>			
Resources to Support Staff Training in Evaluation <sup>a</sup>	9	2	2

*a The N= 13, including: Agriculture, Education, EPA, HHS, HUD, Interior, Justice, Labor, NASA, SSA, Transportation, USAID and VA.*

*b The N= 11, including: Labor: Employment and Training Administration; Justice: Bureau of Prison; and National Institute of Justice; Agriculture: Food and Nutrition Service; USAID; Education: Office of Student Financial Assistance; HHS: Health Resources and Service Administration and Center for Disease Control and Prevention; and Transportation: NHTSA, Maritime Administration and RSPA.*

- Building evaluation skills and knowledge is necessary; and
- Line managers’ knowledge needs to be enhanced about the right kinds of evaluative questions to raise about programmatic performance.

These themes are amplified below.

### **Finding Evaluation Capacity**

Respondents in this study presumably are as familiar as any staff with the level of evaluation capacity in their agencies, yet most were hesitant to offer global judgments. When asked to count the number of evaluators in central policy and planning offices or in counterpart offices in operating units, many were careful to couch their responses. Some operating units contacted are extremely research-oriented, such as HHS’ Centers for Disease Control and DOT’s Research and Special Program Administration, and respondents found it difficult to differentiate evaluators from other analytical staff. In some operating units, respondents were unsure whom to count as evaluators because “it may not be in their job description.” In several central budget and planning offices, the respondents identified

specific staff as evaluators, but then noted that “they do not really evaluate programs.” In general, agencies did not have a defined set of criteria or a qualifications statement for what skills are needed to be an “evaluator.”

Clarity on what constitutes evaluation capacity varied greatly across respondents. In some cases, respondents readily volunteered that program evaluation has recently been revitalized in their agency, acknowledging that this has been due in large part to GPRA. For example, the Central Planning Office of Veterans Affairs reopened an evaluation unit two years ago, after closing it eight years before. The Office of Inspector General at the Environmental Protection Agency (EPA) obtained approval to open a Program Evaluation unit in February of 2000 with 15 new full-time equivalent (FTE) evaluators. New hires in evaluation shops were not apparent across all the agencies surveyed, but there was certainly evidence that there have been some new hires dedicated to evaluation work — even if what these staff will actually do on the job may not be clear. Several respondents mentioned obstacles in federal hiring procedures that make it difficult to hire staff with needed skills.



## Building Evaluation Skills and Knowledge

Funding to provide training on evaluation for staff in policy and evaluation offices and for line managers is critical, yet it was viewed as inadequate across virtually all of the agencies contacted. Most respondents volunteered that personnel in both staff and line offices need more evaluation knowledge and skills, but that training in this area is typically not viewed as a high priority. There has been a flurry of workshops on GPRA implementation and basic performance measurement concepts offered to agency staff directly involved in implementing GPRA, but this training has been tightly focused on immediate responsiveness to GPRA and has typically not been offered to line managers across the agencies.

Some respondents volunteered that a carrot and stick approach to enhancing evaluation knowledge among staff will be necessary. They suggest that more funding for training is necessary. However, they believe that until staff sees evidence that using evaluation to support performance measurement and management brings positive consequences, initiatives to strengthen evaluation skills and knowledge will not be a priority. The incentives must come from top leadership in the agencies, along with the resources.

## Asking the Right Evaluative Questions about Performance

Building evaluation capacity to effectively support performance management in agencies extends to expanding the knowledge and skills held by line managers as well as the priority given to evaluation in the organizational culture. Virtually all respondents recognized the challenges placed upon line managers to ask good, targeted questions about the performance data they are now collecting and/or reviewing. While many respondents were hesitant to identify specific aspects of program evaluation knowledge needed by line managers, they acknowledged a need to prepare managers to apply standards to these new tasks.

Program managers often serve as developers and monitors of their agency's contracts for evaluation studies, but lack the technical background needed

### Department of Health and Human Services (HHS) Capacity Initiatives

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) at HHS is devoting resources to two new initiatives intended to improve evaluation capacity across the operating units:

1. A pool of individual and small business consultants with evaluation expertise has been established via a task-order contracting mechanism to be on call to assist program staff across HHS with developing performance measures, designing evaluation studies that may be contracted out, or any other in-house evaluative work.
2. ASPE contracted for a training needs assessment to be undertaken across the department to identify the specific evaluation skills needed by contract managers in the operating agencies.

for this role. As one respondent aptly volunteered: "The key is to make the program managers more skilled at asking good, hard questions, plus being really good, critical consumers of evaluation work. Managers need to think more like evaluators and view program evaluation as a useful tool."

GPRA has required agencies to measure performance, but the law cannot make managers use the performance data to improve their programs.

The requirements to measure and report on program achievements have pushed managers across most agencies to think about the link between what they do and their performance measures, as well as the quality of the data they receive. While "thinking like an evaluator" about the validity and reliability of data is not new to managers in research-oriented agencies, it is a new task for many managers in other locations.

Pressures from GPRA requirements and agency executives to measure program results or outcomes as well as workloads have also increased pressures for program managers to think critically about the

logic underlying their programs. Systematically modeling programs to identify what to measure requires evaluation knowhow, particularly the use of tools for conceptual development.

Sensitizing program managers to the usefulness of evaluation thinking and tools may require additional training. It definitely requires an organizational culture that supports managers who take the time and effort to systematically and empirically assess their programs.

Successful evaluation projects that inform managers' efforts to measure performance build a more supportive culture for evaluation. Respondents across the Department of Transportation mentioned the value of a recently completed evaluation study on the transportation of hazardous materials (Hazmat) that brought together staff from all its agencies concerned with various transportation modes. The Hazmat evaluation was well received by managers across DOT for a variety of reasons: it demonstrated cooperation across otherwise quite separate operating units; it was designed and overseen by a team of representatives from all the modal agencies working alongside a very supportive Inspector General; it provided useful information to assist managers in improving performance measures and targets; and it provided timely, clear, useful recommendations. Respondents at DOT highlight the impact that the Hazmat evaluation had upon improving the general posture toward evaluation thinking among managers across DOT.

The benefits of such positive evaluation experiences for improving the organizational culture regarding evaluation are invaluable. Positive experiences where the insight offered to managers clearly outweighs the costs imposed upon them are especially effective in building support for evaluation and for enhancing evaluation capacity.

# Organizational Opportunities for Integrating Program Evaluation and Performance Management

As a new administration moves in, agencies face both challenges and opportunities that should be addressed by federal executives. Our respondents emphasized seven major organizational strategies that can be used to support the integration of program evaluation into performance management:

- Secure support from top administrators;
- Institutionalize continuous performance improvement strategies;
- Use a variety of coordinating processes to bring skilled staff together;
- Clarify the roles of key players;
- Work proactively with Congress;
- Build upon information technology to support performance management; and
- Take advantage of initial groundwork institutionalizing performance measurement.

This section briefly describes each of these strategies, with illustrations from the agencies. With the complexity and diversity of operations among the

agencies included in this study, we did not find any single key to more effective use of program evaluation. Just as there are many ways in which the diversity of program evaluation methods can be used to support improved program management, so are there many organizational approaches that support these steps toward performance-based management.

## Secure Support from Top Administrators

Agency respondents were nearly universal in their emphasis on the key roles of top executives. Often a department's deputy secretary has played a major role in obtaining the buy-in of agency administrators by becoming a "champion" for the use of data.

As one federal manager explained: "Starting with the FY 2000 Plan, the deputy secretary took charge. He provides leadership to resolve issues and meet deadlines. He adjudicates major disputes. When managers explain that 'We can't possibly control that situation!' [affecting an indicator], he urges them to use the outcomes more: 'You should

try to meet that indicator.’ In this agency, the Office of the Deputy Secretary is centralized over all the bureaus and provides a centralizing logic for program development.”

Federal departments differ in their extent of centralized versus decentralized evaluation offices and staff. Some respondents described their agency as having program “stovepipes,” with each of its component parts nearly independent, and the departmental Secretary’s office as a “holding company” rather than an active manager of its agency operations. For example, agencies such as the Federal Bureau of Investigation and the Immigration and Naturalization Service have high autonomy within the Department of Justice, as do the operating agencies within the Department of Health and Human Services, such as the Centers for Disease Control and the National Institutes of Health. These “stovepipe” organizations tend not to have a large central evaluation office, and their extent of evaluation use is very mixed, depending on their component agencies’ administrative support for it. The absence of a directive for centralized management means that executive support for evaluation must be present in each component agency in order to strengthen the links to departmental performance management.

In other departments, a centralized evaluation office provides information and guidance on evaluative functions across the agency, such as the Planning and Evaluation Service for the Department of Education. Such central offices were present in only three of the 13 departments contacted, with three others developing this type of office or having some central capacity. These central offices often provide skilled analytic staff for the production of GPRA plans and reports. Respondents in these departments could cite more examples of how they are using program evaluation and other systematic data for GPRA documents and the improvement of program management than did respondents in decentralized departments. *If their central administrators took actions to support data-based management*, then the central evaluation offices seemed to be more involved in using evaluation to support the products and processes for GPRA. In several cases, a new administrator provided the impetus for

increased attention to the data available about program performance, which then cascaded through the agency.

## **Institutionalize Continuous Performance Improvement Strategies**

Many agencies are evolving toward processes that promote continuous improvement in both the use of various types of data to improve program outcomes and in the content of their performance measures. Several respondents characterized these changes as moving toward a “learning organization.” In the first several years of the GPRA requirements, many agencies produced their documents primarily because they were required. Gradually, their managers have begun to see the usefulness of the data. As one observed, “We are now going toward a problem solving mode in trying to solve the problems identified by the performance measures.”

### **Capturing the Interest of Top Administrators**

Respondents mentioned several strategies to increase the interest of top administrators in the use of evaluative data and thus increase their support for it:

- Present data in a way that gets their interest; good graphics are especially useful.
- Summarize information on their pressing policy questions into “quick and easy answers.”
- Provide visibility by asking them to address workshops or other training sessions.
- Build performance reporting into policy initiatives they support.
- Use program “failures” or crises as opportunities for program evaluation to diagnose the source of the problems and show its usefulness.

### **Continuous Improvement at the Department of Justice's Bureau of Prisons**

The Bureau of Prisons uses 20 indicators in its Key Executive Indicators System, which its executive staff examine quarterly. These include intakes and outgoes, population diversity, assault rates, disciplinary actions, educational program enrollment, etc. They look at comparisons among prisons and over time, and if changes occur in one prison but not others, they ask why. The data may be used to trigger a more intensive evaluation, using inside staff experienced with the prison system.

These agencies usually have regular management reviews of their performance data at least once per quarter, not just annually for the GPRA report. They often have many more indicators and data sources for each program than they can report in the Annual Performance Plan or Report, in order to track important program delivery processes and their outcomes. Several mentioned the need to obtain data from program evaluations more quickly than is traditionally possible with contracting mechanisms, such as telephone surveys of consumers or a rapid field appraisal. They emphasize that the involvement and support of top administrators is essential for continuous improvement strategies to be implemented.

The time frames for developing and using valid indicators can be extensive. Many agencies reported they are now changing the indicators that they proposed in their early GPRA Plans, as they put more emphasis on outcomes as well as outputs, obtain feedback on the deficiencies in initial indicators, and develop or obtain access to improved data sources. These continuous improvement processes also depend on team building among the staff involved to make sure that all understand the purposes behind the requirements and to develop the "buy-in" needed among all team members.

In some agencies, the collection and use of information to support program management has long precedents, which has facilitated the availability of

measurement data for GPRA and for continuous improvement. For example, an analyst at the U.S. Agency for International Development noted that country-level mission staff for demography and health are quite adept in using indicators that have been available for some time: "The level of sophistication in using indicators depends on how long we have had program initiatives in that content area. We started earliest with indicators for population and reproductive health, in the 1960s and '70s, then child health in the '80s, and HIV/AIDS and maternal health in the '90s. We started working on infectious disease in 1997, and still don't have good indicators for this area."

### **Use a Variety of Coordinating Processes**

Formal and informal task forces or committees can be used to bring evaluators together with other types of staff to help integrate evaluation and performance management. Most agencies use some type of formal task force or work group charged with putting together performance documents, usually with representatives from each of their operational divisions. Sometimes these representatives are evaluators, but this is not a common requirement. In addition to program staff and evaluators, the work groups often include staff from the budget office, the office of the chief financial officer, the inspector general's office, and scientific support offices, as well as policy development offices. Several agencies have a cascading series of groups, such as a Strategic and Performance Planning Work Group reporting to the top-level Management Review Council in the Department of Labor. These dual arrangements help emphasize high-level interest in the process, while using a task group of mid-level staff to coordinate the more detailed production work among component agencies.

These groups were often begun as temporary structures to meet GPRA requirements, but in several agencies they are now evolving into more permanent coordinating bodies to address issues of strategic planning and evaluation that cut across the department's agencies. For example, both EPA and the Department of Transportation have started a Program Evaluation Council or Network to coordinate evaluation and its relationships to performance measures. In some cases, training or

technical assistance is provided to members of the work groups, with the intention to achieve more stability and experience in these processes. Some of these working groups are focused on a specific topic or issue that needs cross-agency coordination, such as “litigation” within the Department of Justice or transportation of hazardous materials for the Department of Transportation.

Informal relationships are viewed as just as important as the formal processes in many agencies. Respondents stressed the importance of frequent telephone calls, visiting each others’ offices, “hand-holding with the program managers,” and informal conversations in identifying relevant measurement tools and prior evaluations that could enhance a performance report. One of the unanticipated consequences of GPRA may be the stimulus it provides for staff from different parts of an agency, and in different functions, to learn about each others’ roles and about the program strategies of related agencies. For example, in planning for a new cardiac care evaluation, the VA brought together a work group from relevant offices across the agency, including evaluators and program staff, to develop eight focused research questions for the study.

### **Clarify the Roles of Key Players**

As in all efforts to change organizational processes and cultures, the roles and skills of individuals are central to success. Bringing together the diversity of potential evaluative data with the new requirements for GPRA has mandated that people from diverse offices work together more closely and in new ways. Performance management involves a wide variety of people who may not have worked together before, including budget officers, policy developers, program managers, and agency executives, as well as program evaluators and analysts, who often supply the performance data. The intended roles of each type of staff need to be clearly articulated to utilize the skills of all and avoid conflict or duplication of effort.

Performance management may demand more change from program managers than from other types of staff. Often for the first time, they are being asked to think in terms of outcomes for their programs, and to explain program performance mea-

asures to agency executives or Congress. Many respondents noted that managers are the ones on the front line if results do not meet targets. A sampling of their comments: “It is sometimes scary for them.” “Managers may avoid the risk of negative information from evaluation.” “How can they put into place an action plan to improve results for an indicator not under federal control?” “Fear of the unknown is high among program managers when they don’t know how the data will be used.” Positive incentives are needed for managers who obtain and review data about their programs, rather than punishment for not meeting “targets.” Agency executives may need to work with line managers to help them align feasible program strategies with realistic performance measures.

One impetus for managerial change is having the opportunity to examine real data about their programs, a stimulus noted by several interviewees. As one of them put it:

Having the data is a driving factor. Now program managers are finally getting it when they see their data. They are asking good questions about “success” and meeting the targets. What are the comparisons between their “predicted” targets and what the data results showed? They are also thinking more about the measures and their meaning. Some programs within the agency have been embarrassed when other programs have data for performance measures and they don’t. They are asked, if (other programs) can do it, why can’t your program do performance measures?

Similarly, when agencies produce evaluations that a manager finds useful and helpful, the administrators of other programs may ask, “Why aren’t you evaluating my program?”

Attribution for “results” is a key issue for many managers, who recognize that their programs are small pieces in much larger puzzles of federal, state, and local processes. The term “results” implies that the outcome measures are within the control of the federal agency proposing that indicator, but many federal program managers believe they do not control these results for their own pro-

grams. As one administrator explained: “It takes research to find out what strategies are needed and to link program performance and outcomes. Now program managers must make strategy decisions on how to manage their programs and often don’t have data to guide them. Congress and external audiences act as if these relationships are well understood, but they are not.”

Attributing results to program interventions and assessing the linkages between program outputs and outcomes are challenges that can be addressed by program evaluation methods. But the application of these methods requires time, evaluation expertise, and resources to do studies. As indicated by our prior discussion of capacity issues, many agencies do not have or have not budgeted for the resources to do these evaluation studies.

Further, many agencies work in partnership with states, local governments, or other entities, and obtaining performance data from them involves detailed negotiations to obtain good quality, comparable data. Under block grants, there has been limited accountability, and were that to change, local officials may protest, “Just give us the money, and leave us alone!” The new era of accountability also requires a change in the culture of state and local relationships with federal officials. In order to manage the programs to achieve intended results and for GPRA reporting, federal staff need good quality data from their partners and grantees. Yet it takes substantial time, often several years, to work through the problems of indicator definitions and data collection strategies when multiple partners are involved.

## **Work Proactively with Congress**

Timing is especially relevant for federal executives currently attempting to integrate program evaluation and performance measurement given changes in Congress and the presidential administration. Anticipated changes in the composition of congressional appropriations and authorization committees and in agency leadership may induce anxiety for some. But, somewhat surprisingly, the majority of respondents did not fear that momentum in their analytical accomplishments would be lost due to changes in political leadership,

because good performance data will be needed by any new administration.

Officials interviewed were asked whether they have found the relative emphasis in implementing GPRA to be on accountability, specifically reporting to Congress, or on program improvement. The majority responded that so far the emphasis within the agencies has been on accountability, but with a growing focus on using data for program improvement.

The first full cycle of GPRA planning and reporting has been a learning process for all parties. Most agencies started with far too many performance measures, then observed that their appropriations committees were not highly interested anyway. Many agencies are now using a dual strategy of reporting a few broad measures in GPRA accountability documents, but collecting many more indicators for program-specific management and improvement.

Working with Congress to move agencies forward in performance management continues to present a challenge. The role of Congress in encouraging the implementation of GPRA has been sporadic and somewhat confusing for agency staff. Many agency respondents were concerned that if Congress uses GPRA reporting documents in a “heavy-handed way,” it will undermine their efforts to improve data and to use it for performance management. “Fear of the unknown is high among managers when they don’t know how the data will be used,” said one.

Some congressional members have shown an interest in seeing performance information, but most appropriations committees reportedly have worked through the budget as if there were no GPRA. A few appropriations committees have been micro-managing agencies in reviewing their performance plans, but they are the exception, rather than the rule. There have not been consistent signals coming from the appropriations committees on how performance reports will be used in the budget process. And as one respondent noted, generally speaking “it is only the good government types on the Hill who care about GPRA, not appropriations committees.”

Perhaps the best way to summarize the views about Congress’ role in using program evaluation to sup-

port GPRA implementation is uncertainty about the extent of congressional support and extreme variability across the committees. HHS has a legislated one percent set-aside for evaluations of public health service programs. The respondents believed that such legislative budget support is absolutely essential. They pointed out that if Congress asks for systematic evaluation of a program, it gets done, but without such direct requests, it is difficult to include such expenditures in their budgets. If GPRA serves to accustom appropriations committees to receiving useful performance data, it is likely that the learning process will help raise evaluation-type questions. Then, congressional requests will likely increase rather than decrease. In any case, proactive action on the part of executives to focus committees on agency performance is preferable than waiting for Congress to lead.

## **Build Upon Information Technology to Support Performance Management**

Many respondents volunteered that the use of information technology in management information systems supports both performance measurement and further evaluation. It opens opportunities for ongoing evaluation for program improvement when the technology facilitates more frequent submission of data. MIS systems for multi-state programs open opportunities for data mining, such as disaggregation to address significant questions about targeting services and the variability in outcomes among sites. With the accumulation of data over time from multiple participants and partners, rigorous analysis of the impacts of program changes or variations is often possible, using causal modeling.

Further, use of graphics capabilities in modern analytic software helps to communicate evaluation and performance results in “user-friendly” formats. Sophisticated use of Geographic Information Systems at the Department of Justice, and multi-purpose management information systems at the Department of Labor’s Employment and Training Administration, illustrate ways that information technology opens opportunities for managers to present their data clearly and to ask more sophisticated analytical questions about programs and their

performance. After their successful maneuvers to combat the “Y2K bug,” information technology staff may also be available to redirect their MIS systems toward greater support of performance measurement and evaluation.

## **Take Advantage of Initial Groundwork for Institutionalizing Performance Measurement**

Federal executives should build on the momentum already achieved in institutionalizing performance management. While the responses to GPRA among agencies are somewhat uneven, it is still early and much learning has already taken place. The general tone of respondents was very optimistic about the potential for evidence-based performance management in their agencies. Changes are occurring in the ways that managers operate when they have performance data available. For example, managers across agencies are talking about “verification and validation” of data because they know that quality performance measurement and reporting is now business as usual. The respondents noted that now, finally, managers throughout their agencies recognize that GPRA is not simply a paperwork exercise that will go away soon, as have many other government reforms attempted in the past.

Thus far, implementation of GPRA has initiated the mechanisms to support performance management. Performance is being measured and reported, and program strategies are being re-assessed as performance results are reviewed. Initial efforts at measurement have pushed managers to ask good questions about how relevant the measures are to what they do and how good the data actually are. This is progress. In fact, the tools are in place to permit the new leadership taking over in January 2001 to move ahead with performance-based management — if they so choose.

The arrival of new leadership presents an important opportunity for evaluation capacity to be expanded and utilized more effectively to support performance management. In many agencies, much painful deliberation and “heavy lifting” have already taken place regarding GPRA. In some, resources are being allocated to build performance



information systems and to support some of the needed analytical work. Drafters of future Annual Performance Reports have a distinct advantage as they learn from the early performance reporting experiences. False starts and misplaced optimism about what is achievable have occurred, but agency staff have made great progress along the learning curve. The new leaders will find an “older and wiser” GPRA staff awaiting them with better and more informed expertise about how best to use performance measurement systems to move their agencies forward. The most fortunate leaders are those coming to take the helm from top management teams that fared especially well in the first round of GPRA. Much trial and error learning has taken place across the federal government. This experience lays a strong foundation for the new agency executives who take over while the momentum is moving forward, as they build upon the performance systems initiated in the early GPRA work.

# Conclusions and Recommendations

Federal agencies are rising to the analytical challenges thrust upon them by GPRA through effective use of talented staff who have tried to cope with inadequate evaluation capacity through resourceful maneuvers. This report identifies numerous ways in which federal agencies are using program evaluation strategies and tools to support the implementation of GPRA.

A key study finding is that GPRA has raised expectations and demands for the application of evaluation expertise to support systematic analysis of program performance in the federal government. The officials interviewed believed that opportunities for using program evaluation strategies and tools can help performance measurement efforts to improve management and the results of programs. However, in general, evaluation capacity within these agencies is not adequate to contribute effectively to performance measurement and management. Most agencies do not allocate funding to support an adequate number of evaluation staff, nor provide the training needed for either the analytical staff already there or the line managers who are being asked to cope with new analytical challenges.

Most respondents recognized the need for program evaluation strategies and tools to support ongoing performance management. Many examples were described that demonstrate the diversity of ways that evaluation tools are being used or could be

used to support this work. Yet most agencies were using evaluation only tangentially, and their staff acknowledged that evaluation and performance measurement are not yet integrated for many of their programs.

This study developed a conceptual framework to help categorize potential uses of evaluation and to encourage fuller use of this potential as agencies plan for future evaluation. The framework suggests that evaluation can be marshaled to support four stages in agencies' growth toward performance management: strategic and program planning; improving program delivery; accountability to external bodies, such as Congress; and attributing results to the agency's programs. These stages are of course cyclical, and evaluation for each stage can contribute feedback to earlier processes. Four major types of evaluation methods form the rows of the framework: conceptual development tools, such as logic modeling; data collection and quality control; the use of statistical and management information systems; and the use of process and impact evaluation studies. The cells of the framework in Table 3 (page 13) illustrate the great variety of ways that evaluative data could enhance performance management. With broader thinking and planning for the full potential range of evaluation uses, executives can stimulate their agencies to achieve much greater benefit from their existing evaluation resources for improved performance management.

Federal officials involved in GPRA implementation who were surveyed for this study identified a variety of very useful processes they employ to integrate program evaluation and performance measurement. They highlighted the need to receive demonstrated support for this effort from top administrators to create central evaluation capacity and to “champion” their efforts. Several are using “continuous improvement strategies” with at least quarterly reviews of performance data by several levels of management. These include both programmatic and executive level review of detailed performance data and the iterative development of improved measures over time. Good coordinating structures were seen as vital to this work, with task forces evolving into permanent intra-agency committees to achieve stability and continuity, but with continued emphasis on informal relationships among key staff in diverse locations.

The roles of many types of staff contribute to these evolving processes, and clarity of their “division of labor” among offices is needed. Operating-level program managers are likely to be most affected by GPRA reporting requirements, and most anxious about how GPRA will affect them. Evaluators and other analytic staff can relieve these fears by helping them to interpret and use performance data for improving their results, and by applying evaluation techniques to the key concerns about attributing “results” to their programs.

The respondents also identified a number of remaining challenges and opportunities they face as they move into further cycles of performance planning and reporting under GPRA. They are concerned about the uncertain support and mixed signals they receive about performance measurement from Congress, and insufficient resources for strengthening evaluation capacity within their agencies. They recognize that line managers need some retooling in this era of performance-oriented management to help them think more like evaluators as they collect, review, and use performance data, but that this will require resources and positive incentives. On the other hand, they voiced optimism about the opportunities presented by fuller use of information technology and the forward momentum for institutionalizing performance measurement and management in their agencies. They believe the groundwork has been laid and

that managers are moving forward on the learning curve to present inviting opportunities for the new leadership arriving in January of 2001.

A variety of recommendations are offered here to improve the synergistic relationship between program evaluation and performance measurement to foster evidence-based management efforts in the federal government. Improving evaluation and analytical capacity is a key concern of the respondents here. They described many examples showing how ingenious staff leverage the resources available to use evaluative strategies and tools in creative ways. In response, these examples can inform and persuade managers of the potential for more effective use of evaluative information to support performance management. The bottom line is that strengthening evaluation capacity and use will enhance the likelihood that the performance measurement and management framework being institutionalized across the federal government will result in both improved program management and desired results.

## Recommendations

Clearly, many opportunities are available for enlightened agency executives to utilize existing and evolving evaluation strategies and tools to support their agency’s performance measurement efforts. The recommendations that follow identify several actions for executives, evaluators, and agency managers to bridge the gaps between many agencies’ current practices in collecting and using performance data, and potential contributions from the broad field of evaluation.

### Make Fuller Use of Program Evaluation Tools and Skills

#### ***1. Recognize the diversity of ways that evaluation tools can support performance management, and plan for a broad array of evaluative functions.***

By recognizing evaluation as an ongoing function encompassing many opportunities, agencies can support each stage of their work toward full performance management. This study identifies a wide array of ways in which performance improvement and accountability can be enhanced both by fuller use of regularly collected performance data and by complementary discrete evaluative studies.

**2. Conduct an inventory of the evaluation skills of the agency staff directly involved in planning performance measures and reporting them.**

A good match between the evaluation skills needed and the available skills of those contributing to performance reports is likely to improve report quality. We observed that several agencies which integrated staff with evaluation skills into their GPRA production teams tended to receive high ratings for their GPRA efforts, such as the Departments of Education and Transportation. Improving this match may involve several options: 1) transferring staff skilled in evaluation into the offices most closely involved in performance measurement efforts; 2) more training on evaluation for existing staff who contribute to performance planning and reporting; and 3) more structured coordination of work conducted jointly by evaluators and the performance office staff. In agencies where the numbers of evaluators have been substantially reduced by personnel cuts, hiring more staff with evaluation skills may be needed.

**3. Seek out staff with evaluation skills when implementing many of the steps in performance management.**

Including evaluators and others with analytic skills as key members of GPRA work groups is likely to integrate these perspectives with those from other backgrounds. The field of evaluation is now much broader than its earlier focus on methods for rigorous impact evaluation, but many agency administrators do not have a full view of the potential roles for evaluators. Evaluators may need to “sell” the usefulness of their areas of expertise to managers who are skeptical of its applicability to performance measures. At the same time, evaluators need to be supportive of agencies’ initial efforts in performance measurement, which may not meet ideal standards for data quality.

**4. Provide more explicit guidance for program officers who oversee contractors conducting program evaluations.**

Many of those interviewed suggested that more, not less, program evaluation work will be conducted in this period of performance-based government, and that large evaluation studies will continue to be contracted out. To ensure that con-

tracted evaluations meet expectations for relevance and quality, those in charge of overseeing these contracts need sufficient knowledge to construct appropriate statements of work, to ask the right technical questions of the contractors before and while the contracted work is undertaken, and to review the quality of contractors’ draft products (Behn and Kant, 1999).

**Leverage Capacity for Implementing Performance-Based Management**

**5. Persuade new political appointees to allocate resources for building greater evaluation capacity.**

The unambiguous message from respondents is that training and new hires are required to bring evaluation capacity in line with the analytical demands placed upon the agencies now and in the immediate future. Agencies that have developed central evaluation units and/or that have legislated set-aside budgets for evaluation have better track records for meeting GPRA requirements.

**6. Assess and enhance, if necessary, the evaluation skills of the Office of Inspector General (OIG) staff involved in performance data development and auditing.**

Those auditing data quality need to be cognizant of the relevant quality standards and methods for the content area of the performance data. Again, a good match between the measurement-related tasks and the appropriate skills for OIG staff might be improved through one of the methods noted above. The roles of OIG staff in auditing data quality may need ongoing coordination with the work of evaluators in data design for performance measures, via the use of a performance measures work group or another joint planning strategy.

**7. Search out agency resources that might be re-allocated to support using data for performance management.**

Relevant analytic skills may be located in units other than those labeled as “evaluation,” such as in statistical sections, research units, or information technology groups. Again, a broader view of evaluative functions and skills can bring to bear agency data and other resources that might make greater contributions to performance measurement.

## **Integrate Program Evaluation and Performance Management**

### ***8. Institute and support ongoing teams that bring together evaluators from technical offices, program management, OIG staff, and performance planning and reporting staff.***

These groups should be used to transfer knowledge throughout the agency about feasible uses of evaluation methods, as well as performance “success stories.” For example, a Program Evaluation Council has been instituted at the Department of Transportation that provides many opportunities for improved communication among the multiple agencies that make up DOT.

### ***9. Foster basic performance measurement and evaluation skills as important managerial competencies for all line managers.***

At a minimum, program managers should be versed in the basics of program logic modeling, measurement development, descriptive data presentations, and data validity and reliability. The Department of Education is requiring program managers to attest to the quality of the performance data they submit about their programs, using quality criteria developed via a participative process, or to submit plans for data improvement if currently available data are known to be deficient. The department is also providing training and consultation for the managers to help them meet the quality standards.

### ***10. Plan for continuous improvement in evaluation and performance measurement, building on the groundwork laid in the initial rounds of GPRA work.***

Processes for building good data systems and evaluation findings often require a considerable period of time, even five to 10 years. But substantial progress has been made in many agencies in their initial steps toward meeting GPRA requirements. Agencies should assume that this work toward performance-based management will need to be continuous in order to achieve meaningful improvement in program results, rather than a temporary addition to “business as usual.”

### ***11. Publicly reward managers who obtain and use performance data in decision making, for existing as well as new programs.***

The intended outcomes from many programs require longer time periods than the annual time frames in GPRA reporting, so focusing on annual targets may be tangential to their management. Rather than applauding managers who achieve often artificially set “targets,” executives are more likely to help improve their agency’s programs by supporting managers who allocate resources to obtain a meaningful set of output, outcome, and results measures, and then manage their programs to improve the full “chain of events” affecting their results.

### ***12. Use evaluation findings in appropriate ways to amplify data about results when reporting to Congress on GPRA requirements.***

Often a program’s “performance story” is more complex than the broad measures selected to communicate an agency’s high-level goals. By selectively including more detailed information from evaluation studies that illuminate how these results are achieved, agencies can provide a more convincing picture of how and why their programs help to achieve those results.

# References

- Behn, Robert D. and Peter A. Kant. (1999). "Strategies for Avoiding the Pitfalls of Performance Contracting," *Public Productivity and Management Review*, 22, 4: 470-489.
- Bickman, Leonard (1987). "Using Program Theory in Evaluation," *New Directions for Program Evaluation*, 33, Spring.
- Chen, Huey-Tsyh (1990). *Theory-Driven Evaluations*. (Newbury Park, CA: Sage Publications).
- Cooley, Larry (1994). "Challenges and Opportunities for Performance Measurement at USAID," *USAID Evaluation News*, 1.
- Hatry, Harry (1999). *Performance Measurement: Getting Results*. (Washington, D.C.: The Urban Institute Press).
- Hendricks, Michael (2000). "Attribution: Can We Soothe the Achilles' Heel of Performance Measurement?" presentation at the CAP/GAO/GWU Brown Bag Lunch Services, April 19.
- Hendricks, Michael (1996). "Performance Monitoring," AEA Presentation, 1995; and "Performance Monitoring: How to Measure Effectively the Results of Our Efforts," AEA presentation workshop.
- MacLaughlin, John A. and Gretchan B. Jordan (1999). "Logic Models: A Tool for Telling Your Program's Performance Story," *Evaluation and Program Planning*, 22: 65-72.
- Marquart, Jules M. (1990). "A Pattern-Matching Approach to Link Program Theory and Evaluation Data," in *Advances in Program Theory*, L. Bickman, ed. (New Directions for Program Evaluation, #47) San Francisco: Jossey Bass.
- Mayne, John (1999). "Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly." Paper Published by the Canadian Office of the Auditor General of Canada, June.
- National Academy of Public Administration. (2000) Transition Session on the Results of GPRA, Washington, D.C., October.
- Newcomer, Kathryn E. (2000). "Measuring Government Performance," in *Handbook of Public Management Practice and Reform*, K. Tom Liou, ed. (New York: Marcel Dekker, Inc).
- Newcomer, Kathryn E. (1994). "Opportunities and Incentives for Improving Program Quality: Auditing and Evaluating," *Public Administration Review*, 54.
- Perrin, Edward B., Jane S. Durch, and Susan M. Skillman, Eds. (1999). *Health Performance Measurement in the Public Sector: Principles and Policies for Implementing an Information Network*. (Washington, D.C.: National Academy Press).
- Reynolds, Arthur F. (1998). "Confirmatory Program Evaluation: A Method for Strengthening Causal Inference," *American Journal of Evaluation*, 19, 2: 203-221.

- Scheirer, Mary Ann (1994). "Designing and Using Process Evaluation," in *Handbook of Practical Program Evaluation*, J.S. Wholey, H.P. Hatry, and K.E. Newcomer, eds., (San Francisco: Jossey-Bass).
- Scheirer, Mary Ann (2000). "Getting More 'Bang' for Your Performance Measures 'Bucks,'" *American Journal of Evaluation*, 21, 2: 139-149.
- United Way of America (1996). *Measuring Program Outcomes: A Practical Approach*.
- U.S. General Accounting Office (1998). *Managing for Results: An Agenda to Improve the Usefulness of Agencies' Annual Performance Plans*. (GAO/GGD/AIMD-98-228) Washington, D.C.: USGAO.
- U.S. General Accounting Office (1997). *Managing for Results: Analytic Challenges in Measuring Performance*. (GAO/HEHS/GGD-97-138) Washington, D.C.: USGAO.
- U.S. General Accounting Office (2000). *Managing for Results: Continuing Challenges to Effective GPRA Implementation*. (GAO/T-GGD-00-178) Washington, D.C.: USGAO.
- U.S. General Accounting Office (1998). *Managing for Results: Measuring Program Results That Are Under Limited Federal Control*. (GAO/GGD-99-16) Washington, D.C.: USGAO.
- U.S. General Accounting Office (1999). *Managing for Results: Opportunities for Continued Improvements in Agencies' Performance Plans*. (GAO/GGD/AIMD-99-215) Washington, D.C.: USGAO.
- U.S. General Accounting Office (1998). *Performance Measurement and Evaluation: Definitions and Relationships*. (GAO/GGD-98-26) Washington, D.C.: USGAO.
- U.S. General Accounting Office (1998). *Program Evaluation: Agencies Challenged by New Demand for Information on Program Results*. (GAO/GGD-98-53) Washington, D.C.: USGAO.
- U.S. General Accounting Office (2000). *Program Evaluation: Studies Helped Agencies Measure or Explain Performance*. (GAO/GGD-00-204) Washington, D.C.: USGAO.
- Weiss, Carol H. (1998). *Evaluation (Second Ed.)*. (Upper Saddle River, NJ: Prentice Hall).
- Wholey, Joseph S. (1979). *Evaluation: Promise and Performance*. (Washington, D.C.: The Urban Institute).
- Wholey, Joseph S. (1987). "Evaluability Assessment: Developing Program Theory," in *Using Program Theory in Evaluation*, L. Bickman, ed., (New Directions for Evaluation, #33) San Francisco: Jossey-Bass.
- Wholey, Joseph S. (1999). "Performance-based Measurement: Responding to the Challenges," *Public Productivity and Management Review*, 2, 3: 288-307.

# About the Authors

**Kathryn Newcomer** is Professor and Chair in the Department of Public Administration at the George Washington University, where she teaches public and nonprofit program evaluation, research design, and applied statistics. She conducts research and training for federal and local government agencies on performance measurement and program evaluation, and has done consulting work for the governments of the United Kingdom, Ukraine, and Brazil on performance auditing. She is currently consulting with the Smithsonian Institution to design a performance measurement system.

Dr. Newcomer has published three books, *Improving Government Performance* (1989), *The Handbook of Practical Program Evaluation* (1994), and *Using Performance Measurement to Improve Public and Nonprofit Programs* (1997) (all published by Jossey-Bass), and numerous articles in journals including the *Public Administration Review*. Her current research focuses on the changing role of the Office of Inspector General (OIG) in the federal government, as she updates the OIG profiles she provided in 1992 and 1996. She currently serves on the Advisory Council on Auditing Standards for the General Accounting Office, and is a Fellow of the National Academy of Public Administration. Dr. Newcomer has won two awards for her teaching; in 1996 she was awarded the Peter Vail Excellence in Education Award, and in May 2000 she received the George Washington Award.

Dr. Newcomer earned a B.S. in education and an M.A. in political science from the University of Kansas, and her Ph.D. in political science from the University of Iowa.





**Mary Ann Scheirer** is an independent consultant in program evaluation and performance measurement, as well as a member of the adjunct faculty of the Department of Public Administration, George Washington University. She has more than 20 years of experience conducting program evaluations, with recent emphasis on developing performance measurement systems for government agencies, both in the United States and in several other countries. Her research on assessing program implementation has been published as several books, including *Program Implementation: The Organizational Context* (1981) and *A User's Guide to Program Templates: A New Tool for Evaluating Program Content* (1996), as well as in the *Handbook for Practical Program Evaluation* (1994) and in numerous journals.



Dr. Scheirer has worked with many government agencies, including the National Institutes of Health, the Department of Education, the General Accounting Office, the Substance Abuse and Mental Health Administration, and the Health Resources and Services Administration. She served as chair of the Evaluation Review Panel of the U.S. Public Health Service Review of Health Program Evaluations in 1994, 1995, and 1996. She was elected to the boards of directors of the American Evaluation Association and the Consumer Health Foundation, and as president and treasurer of the Washington Evaluators.

Dr. Scheirer's professional training is in applied sociology, including a Ph.D. in sociology and research methods from Cornell University, and a master's degree in public administration from the Graduate School of Public and International Affairs at the University of Pittsburgh.

# Key Contact Information

## To contact the authors:

### **Kathryn E. Newcomer**

Professor and Department Chair  
Department of Public Administration  
School of Business and Public Management  
George Washington University  
302 Monroe Hall  
Washington, D.C. 20052  
(202) 994-6295

e-mail: [newcomer@gwu.edu](mailto:newcomer@gwu.edu)

### **Mary Ann Scheirer**

Consultant in Evaluation and Performance Measurement  
4616 Hillbrook Drive  
Annandale, VA 22003  
(703) 256-8867

e-mail: [MAScheirer@aol.com](mailto:MAScheirer@aol.com)

## ENDOWMENT REPORTS AVAILABLE

### Innovations: Program Delivery

---

**Managing Workfare:** The Case of the Work Experience Program in the New York City Parks Department (June 1999)

Steven Cohen

**New Tools for Improving Government Regulation:** An Assessment of Emissions Trading and Other Market-Based Regulatory Tools (October 1999)

Gary C. Bryner

**Religious Organizations, Anti-Poverty Relief, and Charitable Choice:** A Feasibility Study of Faith-Based Welfare Reform in Mississippi (November 1999)

John P. Bartkowski  
Helen A. Regis

**Business Improvement Districts and Innovative Service Delivery** (November 1999)

Jerry Mitchell

**An Assessment of Brownfield Redevelopment Policies:** The Michigan Experience (November 1999)

Richard C. Hula

**San Diego County's Innovation Program:** Using Competition and a Whole Lot More to Improve Public Services (January 2000)

William B. Eimicke

**Innovation in the Administration of Public Airports** (March 2000)

Scott E. Tarry

**Entrepreneurial Government:** Bureaucrats as Businesspeople (May 2000)

Anne Laurent

**Rethinking U.S. Environmental Protection Policy:** Management Challenges for a New Administration (November 2000)

Dennis A. Rondinelli

### Innovations: Management

---

**Credit Scoring and Loan Scoring:** Tools for Improved Management of Federal Credit Programs (July 1999)

Thomas H. Stanton

**Determining a Level Playing Field for Public-Private Competition** (November 1999)

Lawrence L. Martin

**Using Activity-Based Costing to Manage More Effectively** (January 2000)

Michael H. Granof  
David E. Platt  
Igor Vaysman

**Implementing State Contracts for Social Services:** An Assessment of the Kansas Experience (May 2000)

Jocelyn M. Johnston  
Barbara S. Romzek

**Corporate Strategic Planning in Government:** Lessons from the United States Air Force (November 2000)

Colin Campbell

**The President's Management Council:** An Important Management Innovation (December 2000)

Margaret L. Yao

**Using Evaluation to Support Performance Management:** A Guide for Federal Executives (January 2001)

Kathryn Newcomer  
Mary Ann Scheirer

**Managing for Outcomes:** Milestone Contracting in Oklahoma (January 2001)

Peter Frumkin

### Transforming Organizations

---

**The Importance of Leadership:** The Role of School Principals (September 1999)

Paul Teske  
Mark Schneider

**Leadership for Change:** Case Studies in American Local Government (September 1999)

Robert B. Denhardt  
Janet Vinzant Denhardt

**Managing Decentralized Departments:** The Case of the U.S. Department of Health and Human Services (October 1999)

Beryl A. Radin

**Transforming Government:** The Renewal and Revitalization of the Federal Emergency Management Agency (April 2000)

R. Steven Daniels  
Carolyn L. Clark-Daniels

**Transforming Government:** Creating the New Defense Procurement System (April 2000)

Kimberly A. Harokopus

**Trans-Atlantic Experiences in Health Reform:** The United Kingdom's National Health Service and the United States Veterans Health Administration (May 2000)

Marilyn A. DeLuca

**Transforming Government:** The Revitalization of the Veterans Health Administration (June 2000)

Gary J. Young

**The Challenge of Managing Across Boundaries:** The Case of the Office of the Secretary in the U.S. Department of Health and Human Services (November 2000)

Beryl A. Radin

**A Learning-Based Approach to Leading Change** (December 2000)

Barry Sugarman

**Creating a Culture of Innovation:** Lessons from America's Best Run City (January 2001)

Janet Vinzant Denhardt  
Robert B. Denhardt

### E-Government

---

**Managing Telecommuting in the Federal Government:** An Interim Report (June 2000)

Gina Vega  
Louis Brennan

**Using Virtual Teams to Manage Complex Projects:** A Case Study of the Radioactive Waste Management Project (August 2000)

Samuel M. DeMarie

**The Auction Model:** How the Public Sector Can Leverage the Power of E-Commerce Through Dynamic Pricing (October 2000)

David C. Wyld

**Supercharging the Employment Agency:** An Investigation of the Use of Information and Communication Technology to Improve the Service of State Employment Agencies (December 2000)

Anthony M. Townsend

**Assessing a State's Readiness for Global Electronic Commerce:** Lessons from the Ohio Experience (January 2001)

J. Pari Sabety  
Steven I. Gordon

### Revitalizing the Public Service

---

**Results of the Government Leadership Survey:** A 1999 Survey of Federal Executives (June 1999)

Mark A. Abramson  
Steven A. Clyburn  
Elizabeth Mercier

**Profiles in Excellence:** Conversations with the Best of America's Career Executive Service (November 1999)

Mark W. Huddleston

**Leaders Growing Leaders:** Preparing the Next Generation of Public Service Executives (May 2000)

Ray Blunt

**Reflections on Mobility:** Case Studies of Six Federal Executives (May 2000)

Michael D. Serlin

**Toward a 21st Century Public Service:** Reports from Four Forums (January 2001)

Mark A. Abramson, Editor

## About PricewaterhouseCoopers

The Management Consulting Services practice of PricewaterhouseCoopers helps clients maximize their business performance by integrating strategic change, performance improvement and technology solutions. Through a worldwide network of skills and resources, consultants manage complex projects with global capabilities and local knowledge, from strategy through implementation. PricewaterhouseCoopers ([www.pwcglobal.com](http://www.pwcglobal.com)) is the world's largest professional services organization. Drawing on the knowledge and skills of more than 150,000 people in 150 countries, we help our clients solve complex business problems and measurably enhance their ability to build value, manage risk and improve performance in an Internet-enabled world. PricewaterhouseCoopers refers to the member firms of the worldwide PricewaterhouseCoopers organization.

## For additional information, contact:

**Mark A. Abramson**

Executive Director

The PricewaterhouseCoopers Endowment for  
The Business of Government  
1616 North Fort Myer Drive  
Arlington, VA 22209

(703) 741-1077

fax: (703) 741-1076

e-mail: [endowment@us.pwcglobal.com](mailto:endowment@us.pwcglobal.com)

website: [endowment.pwcglobal.com](http://endowment.pwcglobal.com)

PRICEWATERHOUSECOOPERS 

The PricewaterhouseCoopers Endowment for

## The Business of Government

1616 North Fort Myer Drive  
Arlington, VA 22209-3195

Bulk Rate  
US Postage  
**PAID**  
Permit 1112  
Merrifield, VA