

July 1999

PERFORMANCE PLANS

Selected Approaches for Verification and Validation of Agency Performance Information





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

General Government Division

B-281215

July 30, 1999

The Honorable Fred Thompson
Chairman, Committee on Governmental Affairs
United States Senate

Dear Mr. Chairman:

The Government Performance and Results Act of 1993 (Results Act) seeks to improve the effectiveness, efficiency, and accountability of federal programs by requiring federal agencies to set goals for program performance and to report on annual performance compared with the goals. Annual program goals are to be set out in annual performance plans, and performance against these goals is to be reported in annual performance reports. The first performance reports are to be submitted to the President and Congress no later than March 31, 2000.

In order to credibly report progress toward intended results and to use the information for program management, agencies will need to have sufficiently trustworthy performance information. The Results Act requires agency performance plans to “describe the means to be used to verify and validate measured values” of performance. Verification includes the assessment of data completeness, accuracy, and consistency and related quality control practices. Its purpose is to ensure that the data will be of sufficient quality to document performance and support decision-making. Validation is the assessment of whether the data are appropriate for the performance measure.

In a December 1997 letter to the Director of the Office of Management and Budget, congressional leadership stated that performance plans based on incomplete or inaccurate data would be of little use to Congress or the executive branch. Agencies submitted annual performance plans in spring 1998 and 1999, setting goals for fiscal years 1999 and 2000, respectively. In response to congressional requests, we have reviewed the fiscal year 1999 and 2000 performance plans of the 24 agencies covered by the Chief Financial Officers (CFO) Act. Our analyses of the fiscal year 1999 performance plans concluded that most of the plans reviewed provided limited confidence that agencies’ performance data would be credible.¹ In

¹Managing for Results: An Agenda to Improve the Usefulness of Agencies’ Annual Performance Plans, (GAO/GGD/AIMD-98-228, Sept. 8, 1998).

the report on their assessment of the 1999 performance plans, the House leadership noted that “most agencies lack the reliable data sources and systems needed to develop, validate and verify performance information.”² The report also noted that the problems in performance data were deep-seated and resolving them would take much time and effort. Our assessment of the fiscal year 2000 performance plans identified a continuing lack of confidence in performance information as a major concern. Ultimately, performance plans will not be fully useful to congressional decisionmakers unless and until this key weakness is resolved.³

In this report, as you requested, our objective is to identify reasonable approaches that agencies have proposed or adopted to verify and validate performance information. This report describes these approaches in order to help agency managers select appropriate techniques for assessing, documenting, and improving the quality of their performance data.

Results in Brief

Overall, we found examples illustrating a wide range of possible approaches for increasing the quality, validity, and credibility of performance information. (See app. I for a discussion of how agencies may decide on specific approaches.) These approaches included a variety of senior management actions, agencywide efforts, and specific program manager and technical staff activities. These approaches can be organized into four general strategies, as follows.

Management can seek to improve the quality of performance data by fostering an organizational commitment and capacity for data quality (see app. II). Managers are ultimately responsible for the quality of performance information. We found examples of management communications and actions to encourage the needed coordination, resource allocation, and attention to data quality issues. Reporting efforts to build organizational commitment to obtaining, maintaining, and using good information and to developing the organization’s capacity to do so can help improve the credibility of performance information.

Verification and validation can include assessing the quality of existing performance data (see app. III). Assessments might target specific measures in the performance plan or more broadly assess major

²U.S. Congress, Seeking Honest Information for Better Decisions (<http://freedom.house.gov/results/Implement/Implement4.asp>, June 1998).

³ Managing for Results: Opportunities for Continued Improvements in Agencies’ Performance Plans (GAO/GGD/AIMD-99-215, July 20, 1999).

data systems to identify problems that may affect the use of performance data. In our examples, assessments were conducted internally, built into ongoing work processes and data systems, or involved independent verification and external feedback.

Assessments of data quality are of little value unless agencies are responding to identified data limitations (see app. IV).

Communicating significant data limitations and their implications allows stakeholders to judge the data's credibility for their intended use and to use the data in appropriate ways. In addition to examples of reporting data limitations and their implications in performance plans or other formats, we saw examples of efforts to improve, supplement, or replace existing data.

Building quality into the development of performance data may help prevent future errors and minimize the need to continually fix existing data (see app. V).

Reporting efforts to improve existing data systems or processes can improve the credibility of performance information. We found examples of efforts to build in data quality, including involving stakeholders; providing feedback on data quality problems; and using accepted practices in planning, implementing, and reporting performance data.

Within these general strategies are more specific approaches that agencies may choose to adopt. These specific approaches are listed in figure 1 and discussed in more detail in appendixes II-V of this report.

We identified a wide range of reasonable approaches that agencies can use, where appropriate, to improve the quality, usefulness, and credibility of performance information. How an agency approaches data verification and validation depends on the unique characteristics of its programs, stakeholder concerns, performance measures, and data resources. For example, different approaches may apply to the information collected directly by a federal agency than to that obtained from state sources. Verifying and validating information on client satisfaction may require different approaches than information obtained by direct measurement of environmental conditions, for example.

We expect that agencies will choose from among the approaches described here or will develop different ones to arrive at a systematic strategy suitable to their own situation and performance information sources. Appendix I discusses a number of key questions that can arise when agencies are deciding on the effort to be devoted to verification and

validation, the specific approaches to be adopted, and how to credibly report on their verification and validation efforts.

Figure 1: Menu of Agency Approaches for Verifying and Validating Performance Information

Fostering Organizational Commitment and Capacity for Data Quality

1. Communicate support for quality data.
2. Review organizational capacities and procedures for data collection and use.
3. Facilitate agencywide coordination and cooperation.
4. Assign clear responsibilities for various aspects of the data.
5. Adopt mechanisms that encourage objectivity and independence in collecting and managing data.
6. Provide responsible staff with training and guidance for needed skills and knowledge.

Assessing the Quality of Existing Data

1. Build data quality assessment into normal work processes, including ongoing reviews or inspections.
2. Use software checks and edits of data on computer systems and review their implementation.
3. Use feedback from data users and other stakeholders.
4. Compare with other sources of similar data or program evaluations.
5. Obtain verification by independent parties, including the Office of the Inspector General.

Responding to Data Limitations

1. Report data limitations and their implications for assessing performance.
2. Adjust or supplement problematic data.
3. Use multiple data sources, with offsetting strengths and limitations.
4. Improve the measure by using another source or new methods of measurement.

Building Quality Into the Development of Performance Data

1. Use prior research or analysis to identify data elements that adequately represent the performance to be measured.
2. Gain agreement among internal and external stakeholders about a set of measures that are valid for their intended uses.
3. Plan, document, and implement the details of the data collection and reporting systems.
4. Provide training and quality control supervision for all staff who collect and enter data, especially at local levels.
5. Provide feedback to data collectors on types of errors found by data checks.
6. Use analytic methods and transformations appropriate for the data type and measure being reported.

Because of the need to develop a strategy that meets the unique circumstances of each agency, the framework we present is not a list of requirements or a checklist of steps to follow. Individual agencies should not necessarily be expected to use all of the approaches that we describe, and there may be other approaches that we have not identified.

Scope and Methodology

We conducted our work in six agencies: the Departments of Education, Transportation, and Veterans Affairs and the Environmental Protection Agency, National Science Foundation, and Office of Personnel Management. These agencies were selected after further review of the verification and validation information contained in the 24 annual performance plans we had assessed in 1999. We selected the agencies that we judged would provide a wide range of examples of reasonable verification and validation approaches and represented a variety of performance measurement contexts. These six agencies differed in the extent to which they

- provided internal services to government or to the public,
- conducted and supported scientific research, or
- administered regulatory programs.

They also varied in the extent to which their programs were carried out through the states or delivered directly.

We identified examples of specific verification and validation approaches based on our review of the 1999 and 2000 performance plans and discussions with agency officials. Some of the additional examples identified by agency officials pertain to verifying and validating the performance information used for managing agency programs—not for assessing progress toward performance plan goals. We reviewed agency documents, where available, to confirm our examples and obtain additional detail.

We selected examples that appeared reasonable and useful for other agencies to emulate because they were consistent with accepted professional practice for managing data quality. We found many more examples than are reported here, but we restricted our choices to two or three examples per approach. We included examples to represent different aspects of the approaches being discussed and to provide for a balance among the agencies reviewed.

Where possible, our selection of examples drew on our previous work on the adequacy of the agencies' performance information in specific program

areas. We did not do additional work to assess the adequacy or extent of agencies' implementation of these approaches, the quality of the performance information, or whether the approaches had contributed to improving data quality. Although we identified several specific reasonable approaches that these agencies had in place or planned to implement, our separate assessment of the 24 CFO Act agencies' fiscal year 2000 plans (including the six reviewed for this report) found that none provided full confidence that their performance information would be credible. The performance plans for the six agencies included in this report all provided at least limited confidence in their performance information--that is, they all addressed to varying extents, but not completely,

- their efforts to verify and validate performance data,
- actions to compensate for unavailable or low-quality data, and
- the implications of data limitations for assessing performance.

We developed the framework depicted in figure 1 by analyzing our examples and by reviewing related professional literature. We sought comments on the framework from external professionals and agency officials and incorporated their suggestions where appropriate.

We focused our review on the verification and validation of nonfinancial performance information. Generally, financial performance information that is derived from the same systems that produce financial statement information is subject to the internal control standards, federal financial systems requirements, and accounting standards applicable to federal agencies' financial statement information.

We conducted our work between October 1998 and April 1999 in accordance with generally accepted government auditing standards.

Agency Comments

We did not seek comments on this report because it does not provide an overall assessment of agency data verification and validation efforts. However, we asked officials in each of the six agencies to verify the accuracy of the information presented. We incorporated their clarifications where applicable.

As agreed with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from the date of this letter. At that time, we will send copies to Senator Joseph I. Lieberman, Ranking Minority Member of your Committee; Representative Richard K. Armey, Majority Leader; Representative Dan Burton, Chairman, and Representative Henry A. Waxman, Ranking Minority Member, House

Committee on Government Reform; and Jacob Lew, Director, Office of Management and Budget. We will also make copies available to others on request.

This report was prepared under the direction of Stan Divorski. Mary Ann Scheirer was a key contributor. If you have any questions regarding this report, please contact me or Stan Divorski at (202) 512-7997.

Sincerely yours,

A handwritten signature in black ink that reads "Susan S. Westin". The signature is written in a cursive style with a large initial 'S'.

Susan S. Westin
Associate Director, Advanced Studies
and Evaluation Methodology

Contents

Letter		1
Appendix I		12
Key Questions About	What Are Verification and Validation?	12
Verification and	Why Are Verification and Validation Important?	12
Validation	What Is Data Quality?	13
	How Good Do Data Need to Be?	13
	Does a Given Verification and Validation Approach Apply to All Programs?	14
	Why Attend to Data Quality?	15
	How Can Verification and Validation Efforts Be Clearly Reported?	16
Appendix II		20
Fostering	Communicate Support for Quality Data	20
Organizational	Review Organizational Capacities and Procedures for Data Collection and Use	21
Commitment and	Facilitate Agencywide Coordination and Cooperation	22
Capacity for Data	Assign Clear Responsibilities for Various Aspects of the Data	23
Quality	Adopt Mechanisms That Encourage Objectivity and Independence in Collecting and Managing Data	24
	Provide Responsible Staff With Training and Guidance for Needed Skills and Knowledge	25
	An Example of Agencywide Capacity-building at ED	26
Appendix III		28
Assessing the Quality	Build Data Quality Assessment Into Normal Work Processes, Including Ongoing Reviews or Inspections	28
of Existing Data	Use Software Checks and Edits of Data on Computer Systems and Review Their Implementation	30
	Use Feedback From Data Users and Other Stakeholders	30
	Compare With Other Sources of Similar Data or Program Evaluations	32
	Obtain Verification by Independent Parties, Including the Office of the Inspector General	32
	Consequences From Assessing the Quality of Existing Data	34

<hr/>	
<p>Appendix IV Responding to Data Limitations</p>	<p>Report Data Limitations and Their Implications for Assessing Performance 35</p> <p>Adjust or Supplement Problematic Data 37</p> <p>Use Multiple Data Sources With Offsetting Strengths and Limitations 37</p> <p>Improve the Measure by Using Another Source or New Methods of Measurement 38</p>
<hr/>	
<p>Appendix V Building Quality Into the Development of Performance Data</p>	<p>Use Prior Research or Analysis to Identify Data Elements That Adequately Represent the Performance to Be Measured 40</p> <p>Gain Agreement Among Internal and External Stakeholders About a Set of Measures That Are Valid for Their Intended Use 40</p> <p>Plan, Document, and Implement the Details of the Data Collection and Reporting Systems 41</p> <p>Provide Training and Quality Control Supervision for All Staff Who Collect and Enter Data, Especially at Local Levels 44</p> <p>Provide Feedback to Data Collectors on Types of Errors Found by Data Checks 45</p> <p>Use Analytic Methods and Transformations Appropriate for the Data Type and Measure Being Reported 46</p> <p>An Alternative Approach to Performance Assessment at NSF 46</p>
<hr/>	
<p>Figures</p>	<p>Figure 1: Menu of Agency Approaches for Verifying and Validating Performance Information 4</p> <p>Figure I.1: Sample Presentation Format From ED's Fiscal Year 2000 Performance Plan 18</p> <p>Figure II.1: Approaches to Fostering Organizational Commitment and Capacity 20</p> <p>Figure III.1: Approaches to Assessing the Quality of Existing Data 28</p> <p>Figure IV.1: Approaches for Responding to Data Limitations 35</p> <p>Figure V.1: Approaches to Building Data Quality Into the Development of Performance Data 39</p>

Abbreviations

ANSI	American National Standards Institute
BTS	Bureau of Transportation Statistics
DOT	Department of Transportation
ED	Department of Education
EPA	Environmental Protection Agency
GPRA	Government Performance and Results Act
HEDIS	Health Plan Employer Data and Information Set
ISTEA	Intermodal Surface Transportation Efficiency Act
IPBS	Integrated Performance and Benchmarking System
NCES	National Center for Education Statistics
NCQA	National Committee for Quality Assurance
NHTSA	National Highway Traffic Safety Administration
NSF	National Science Foundation
OMB	Office of Management and Budget
OPM	Office of Personnel Management
RIS	Retirement and Insurance Service
SEA	State educational agencies
SEDCAR	Standards for Education Data Collection and Reporting
VA	Department of Veterans Affairs
VBA	Veterans Benefits Administration
VHA	Veterans Health Administration

Key Questions About Verification and Validation

Many questions can arise when agencies are deciding on the effort that should be devoted to verification and validation and the specific approaches appropriate to their agency and program contexts. These questions can include

- What are verification and validation?
- Why are verification and validation important?
- What is data quality?
- How accurate do data need to be?
- Does a given verification and validation approach apply to all programs?
- Are Results Act requirements the only reasons for attending to data quality?
- How can accountability for verification and validation be clearly reported?

What Are Verification and Validation?

Verification and validation refer to aspects of quality control needed to ensure that users can have confidence in the reported performance information. We define these terms as follows:

- Verification is the assessment of data completeness, accuracy, consistency, timeliness, and related quality control practices.
- Validation is the assessment of whether data are appropriate for the performance measure.

We are not addressing here other aspects of “validity,” such as the appropriateness of the agency’s choice of performance measures in relation to its goals and objectives.¹ Other GAO products discuss issues related to other aspects of validation. For example, our [Results Act: Evaluator’s Guide](#) provides guidance on defining expected performance, and on validly connecting missions, goals, and activities.²

Why Are Verification and Validation Important?

Both verification and validation help to ensure that data of sufficient quality will be available when needed to document performance and support decision-making. To be useful in reporting to Congress on the fulfillment of Results Act requirements and in improving program results, the data must also be “credible,” that is, they must be seen by potential users to be of sufficient quality to be trustworthy.

¹The term “validation” can be used in many different ways, including validation of the appropriateness of the agency’s overall goals and objectives, given the agency’s legislative mandates and mission; performance measures as “validating” the program, e.g., to provide evidence of program results; assessing whether the performance measures chosen by the agency are clearly related to the target objectives; and its more limited use in this report.

²The Results Act: An Evaluator’s Guide to Assessing Agency Annual Performance Plans (GAO/GGD-10.1.2, Apr. 1998).

Reporting validation and verification procedures helps to ensure that data will be credible to potential users. The Office of Management and Budget (OMB) guidance states that “The means used should be sufficiently credible and specific to support the general accuracy and reliability of the performance information . . .”(OMB Circular A-11, sec. 220.13). Attention to “credibility,” in addition to more technical aspects of data quality, requires careful consideration of the needs of the audiences for the information.

What Is Data Quality?

Choices among potential verification and validation approaches involve senior management’s making decisions about data quality. The approaches that agencies use to verify and validate performance data should address key dimensions of data quality. Specific agencies and professional sources have developed data quality criteria specifically relevant to their context and content area. The key dimensions of data quality suggested below were developed for this report to illustrate the types of quality concerns that agencies consider. These are not intended to be exhaustive of all potential quality considerations, nor substituted for agency-developed criteria. Examples of data quality elements are the following:

- **Validity**—the extent to which the data adequately represent actual performance.
- **Completeness**—the extent to which enough of the required data elements are collected from a sufficient portion of the target population or sample.
- **Accuracy**—the extent to which the data are free from significant error.
- **Consistency**—the extent to which data are collected using the same procedures and definitions across collectors and times.
- **Timeliness**—whether data about recent performance are available when needed to improve program management and report to Congress.
- **Ease of use**—how readily intended users can access data, aided by clear data definitions, user-friendly software, and easily used access procedures.

How Good Do Data Need to Be?

There is no easy answer to the question of how good data need to be. No data are perfect. In general, data need to be good enough to document performance and support decision-making. Decisions as to “how good is good enough” may depend on the uses of the data and the consequences of program or policy decisions based on those data. These factors may involve trade-offs among the dimensions of data quality presented above. On the one hand, emphasizing the completeness of a planned data collection effort may reduce its timeliness when data are to be obtained from a large number of independent entities (such as school districts or industrial establishments). On the other hand, seeking to increase timeliness by using a scientific sampling procedure to reduce the number

of entities providing data would reduce the completeness of coverage of entities from which data are collected, but may still provide adequate data for performance measurement.

Different Measures May Require Different Levels of Accuracy

Different levels of accuracy may be needed in different circumstances. For example, audits of financial data, assessments of the extent of air pollution to use in environmental performance measures, and opinion surveys may all require different levels of accuracy. Within these areas, professional judgment plays a role in determining acceptable error levels. In sample surveys, it is recognized that the margin of error “should be dictated by how much error the investigators feel they can live with in reporting survey results.”³

Amount of Change Desired Affects Data Standards

The amount of desired change in performance can also influence the determination of a reasonable data standard. If the amount of error is not sufficiently less than the amount of change targeted, it will not be possible to determine whether change in the measured value is due to error or actual changes in performance.

Does a Given Verification and Validation Approach Apply to All Programs?

Agencies use different approaches to validate and verify performance measures for different types of programs. For example, service delivery assistance programs (e.g., of the Department of Education) or administration of earned benefits (e.g., benefits to federal employees administered by the Office of Personnel Management) might use performance data from a survey of beneficiaries, with data quality criteria derived from appropriate procedures for sample surveys. In contrast, regulatory programs, such as those of the Environmental Protection Agency, may measure targeted pollutants, based on scientifically derived procedures for assessing each pollutant. Determining the most appropriate validation and verification approaches for each type of program is a matter for individual agency diagnosis, analysis, and choice, taking into account stakeholder views, the relevant professional standards, and technical advice.

Useful verification and validation approaches will also vary with the data source being used for the performance measure. For example, if agencies have substantial direct control over data that they generate during their normal operations (e.g., while processing claims for benefits due), agency managers can directly supervise quality control. In contrast, if agency partners, such as state or local grantees, collect the performance data,

³Lu Ann Aday, *Designing and Conducting Health Surveys: A Comprehensive Guide* (San Francisco: Jossey-Bass, 1989), p. 120.

substantial negotiation may be needed to agree on what data elements are feasible to collect and how quality is to be ensured.

Why Attend to Data Quality?

Agencies identified several influences, in addition to the Results Act, that encourage attention to data quality. These influences can be external or internal in origin, and attending to them can help improve programs.

External Factors Can Encourage Data Quality

External influences include external critics, legislative mandates, and the need to comply with professional standards in each program delivery area, such as health care delivery. For example, Department of Transportation (DOT) officials identified several sources that stimulated its efforts to improve data quality. These sources included the Results Act, the National Performance Review, and provisions of the Intermodal Surface Transportation Efficiency Act. This legislation set out explicit data requirements for DOT and created the Bureau of Transportation Statistics with the mission to compile statistics and improve the quality of agency data.

External studies by the National Performance Review and the National Academy of Public Administration identified a need for the Environmental Protection Agency (EPA) to work to improve the quality of environmental data. EPA has also been striving to increase the availability of environmental data to the public and has recently established the expansion of the public's right to know about their environment as a strategic goal. The increased availability of data has brought data quality issues to the surface, and external stakeholders have identified inaccuracies in EPA data.

Data Use Is an Internal Factor That Encourages Data Quality

Agency staff cited increased use of data for program management as an internal influence leading to better data quality. For example, EPA identified the limited operational use of some data as a root cause of errors in the data, as no operational unit was using those data in managing its work. The agency described a variety of potential uses for environmental data, including regulatory compliance, public right-to-know, environmental status and trends reporting, program management, and performance accountability tracking, as called for under the Results Act. To help achieve a better fit between its data and this variety of uses, the agency has undertaken suitability assessments of key data systems for uses other than those originally intended.

Attention to Data Quality Can Help Improve Programs

Agency attention to data quality can help managers increase efficiency or improve operations—for example, by improving the flow of program tasks and the coordination among related programs.

Tracing data flow as a part of assessing the quality of existing data can also lead to improvements in the “business processes” that rely on that data.⁴ In examining the reasons for problems with data quality, EPA identified a lack of correspondence between data systems and overall business process management at the program level as one of the factors. Veterans Benefits Administration (VBA) officials noted that efforts were undertaken to reengineer its business processes to better meet the needs of individual veterans and to improve information about VBA’s business and veterans. These efforts included new procedures to credit benefit processors for their work, changing their focus from an emphasis on “timeliness” to a broader range of work quality criteria in order to improve services to veterans and reduce the potential for distorting measures. These criteria are reflected in a “balanced scorecard” for its performance measures (see app. V).

Validation and verification are not isolated, technical concerns relevant solely to the requirement of the Results Act. Use of data is part of good management. Further, the production of data to inform both business concerns and the public is a fundamental mission of such government agencies as EPA and DOT. Therefore, fostering data quality is fundamental to total agency management. Obtaining agency commitment to and capacity for data that can be verified and validated is a major management issue addressed in appendix II.

How Can Verification and Validation Efforts Be Clearly Reported?

In our reports on agency fiscal year 1999 and 2000 performance plans, we concluded that they provided limited confidence that performance data would be credible, observing that the plans lacked specific information on verification and validation procedures and on data limitations.⁵ Our assessment of the 2000 performance plans noted that most did not identify actions that agencies were undertaking to compensate for the lack of quality data.

Our report on practices that can improve performance plan usefulness to congressional and other decisionmakers identified a number of ways

⁴For a more detailed discussion of the relationship between data quality control and business processes, see Thomas C. Redman, *Data Quality for the Information Age* (Norwood, MA: Artech House, 1996).

⁵GAO/GGD/AIMD-98-228, Sept. 8, 1998 and GAO/GGD/AIMD-99-215, July 20, 1999.

agencies could describe their capacity to gather and use performance information.⁶ These practices include

- identifying internal and external sources for data,
- describing efforts to verify and validate performance data,
- identifying actions to compensate for unavailable or low-quality data, and
- discussing implications of data limitations for assessing performance.

Highlight Verification and Validation Procedures

In addition to these, our current review found examples where agencies enhanced the communication of verification and validation approaches by highlighting them and the data source being verified and validated. We also observed opportunities for agencies to enhance the credibility of their performance plans through more emphasis on verification and validation procedures already in place.

For example, the Department of Education (ED) in its fiscal year 2000 plan used a format that reflects a number of practices to help communicate verification and validation approaches. Figure 2 shows the format ED used to explicitly define each indicator and comment on its background, succinctly describe the implications of a data limitation, briefly present verification and validation information, and identify the data source. A similar format was used by DOT in an appendix to provide information on the data source, verification and validation procedures, and limitations for each measure.

In addition to presenting verification and validation specific to individual measures and data sources, ED and DOT used a separate section to highlight general verification and validation procedures that applied across a number of measures or data sources.

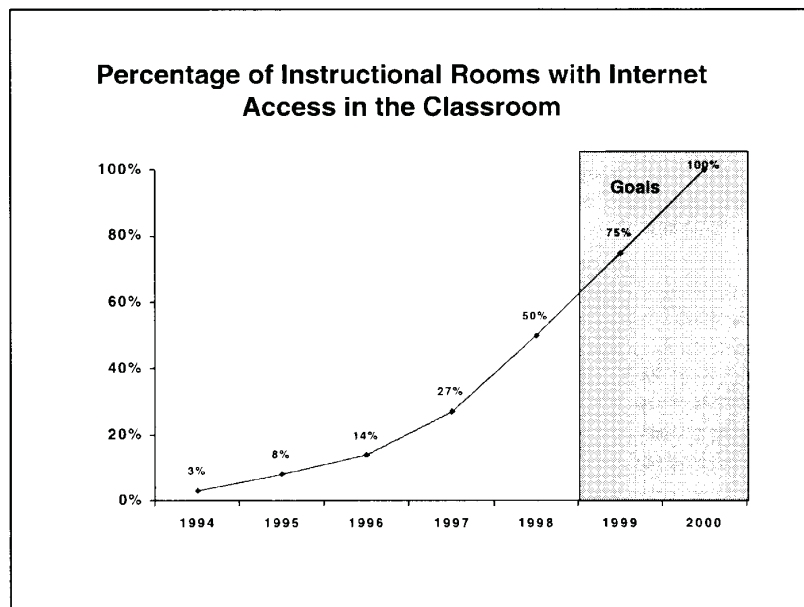
⁶Agency Performance Plans: Examples of Practices That Can Improve Usefulness to Decisionmakers (GAO/GGD/AIMD-99-69, Feb. 26, 1999).

Figure I.1: Sample Presentation Format From ED's Fiscal Year 2000 Performance Plan

Indicator 39. *The percentage of public school instructional rooms connected to the Internet (Information Superhighway) will increase from 14% in 1996 to higher percentages thereafter.*

Indicator background and context. Connections to the Internet make computers versatile and powerful learning tools by introducing students and teachers to new information, people, places, and ideas from around the world to which they might not otherwise be exposed. In 1993, only 3% of instructional rooms were connected to the Internet. By 1997, 27% of class rooms were connected to the Internet (see figure 44).

Figure 44



Limitations of the data. The sample size limits ability to report for subnational units such as individual states, and for subgroups.

Verification/validation of measures: Items have been pretested and used repeatedly. Results are roughly consistent with results from other sources.

Data source(s). National Center for Education Statistic (NCES), Survey of Advanced Telecommunications in U.S. Public Elementary and Secondary Schools, March 1998. Next update: February 1999.

Report Key Approaches
Already in Place

Some agencies' performance plans provide descriptions of planned quality control procedures, without including some ongoing procedures whose description could have increased the credibility of their measures. For example, ED's fiscal year 2000 plan does not describe the extensive quality control procedures already in place for its ongoing national student testing program, the National Assessment of Educational Progress. The Department plans to use measures from this program for several key indicators of major objectives, such as Indicator 2: "Students in high poverty schools will show continuous improvement in achieving proficiency levels comparable to those for the nation." This program is managed by the National Center for Educational Statistics, using credible procedures and expert involvement that could have been summarized in ED's plan.

As another example, the Office of Personnel Management's (OPM) plan does not mention several existing quality control procedures in place for a management information system that will provide a number of indicators for OPM's Retirement and Insurance Service. The plan does briefly mention that verification is undertaken by its Quality Assurance Division, but does not describe approaches used by program management, such as the use of a "physical inventory" to check work processing statistics and accuracy checks on death claims processing.

Fostering Organizational Commitment and Capacity for Data Quality

Obtaining quality performance information is an agencywide management issue, as well as one requiring special attention from technical and program staff. Management needs to create a climate that encourages the needed coordination, resource allocation, and attention to data quality issues that enable improvements in data quality. Several agencies are making efforts to stimulate such a commitment to obtaining, maintaining, and using good information and to developing the organization's capacity to do so. The approaches that agencies are adopting to foster organizational commitment and capacity are shown in figure II.1 and discussed below.

Figure II.1: Approaches to Fostering Organizational Commitment and Capacity

- Communicate support for quality data.
- Review organizational capacities and procedures for data collection and use.
- Facilitate agencywide coordination and cooperation.
- Assign clear responsibilities for various aspects of the data.
- Adopt mechanisms that encourage objectivity and independence in data collection and data management processes.
- Provide responsible staff with training and guidance for needed skills and knowledge.

Communicate Support for Quality Data

Senior agency executives play an important role in fostering program management and staff commitment to data quality. For agency staff and mid-level managers to put priority on data quality, they need to see that senior management values and will use quality performance information for decision-making. We learned from agency officials that data quality is a higher priority when program staff and management see that data will be used for management. Senior executives can provide confidence that they value and will use good quality data by communicating its importance, making data quality an organizational goal, creating a climate of managing for results, and providing technical and financial support.

VBA Senior Management Emphasized the Importance of Accurate Data

For example, in response to an audit by the Department of Veterans Affairs (VA), Office of the Inspector General, senior officials of the Veterans Benefits Administration (VBA) have used presentations and written communications to emphasize to staff the importance of accurate data. The Inspector General concluded that data on the timeliness of processing veterans' claims for benefits were not accurate enough to provide

meaningful measures of VBA's performance. Senior management acknowledged to their staff that data were inaccurate and emphasized the implications of inaccurate data. They also asked staff to undertake reviews to ensure the accuracy of management reports. Management provided staff with a list of unacceptable practices that influence data accuracy and were instructed to stop those practices immediately. Agency officials reported that the communications from senior management resulted in increased attention to data quality and in improvements to data accuracy, such as more accurate recording of the time taken to process compensation claims.

VBA has also moved to foster an organizational commitment to data quality through establishing a related organizational goal in its strategic plan, which is that "VBA's data systems will be reliable, timely, accurate, integrated, honest, and flexible."¹

VHA Holds Managers Accountable for Program Performance

Progress in managing for results also appears to have resulted in greater attention to data quality at the Veterans Health Administration (VHA). One strategy employed by VHA to encourage managers to focus on results was the initiation of a performance contract system. In this system, the Under Secretary for Health negotiates performance agreements with all senior executives in VHA that hold them accountable for quantifiable performance targets. Although these targets do not include ones for data quality, VHA officials told us that assessing managers' performance against them has resulted in greater attention to data quality.

EPA Provides Technical and Financial Support to Improve Data Quality

In another example, Environmental Protection Agency (EPA) management created an initiative that included the provision of technical and financial support for improving data quality. EPA's One-Stop Program is a long-term effort to develop a coherent overall environmental reporting system to address reporting burden and lack of integrated data. EPA will provide technical support and financial assistance to states for developing information management infrastructure and processes, including the adoption of standard data elements.

Review Organizational Capacities and Procedures for Data Collection and Use

Fostering organizational capacity to produce and maintain good quality performance information may require assessing existing organizational capacities and procedures using external or internal reviews. Organizational capacities that might be assessed include the appropriate location of responsibilities for integrating and coordinating data; sufficient staff and expertise to fulfill these responsibilities; appropriate hardware

¹Veterans Benefits Administration, *Roadmap to Excellence: Planning the Journey* (May 29, 1998).

and software; and resources for building, upgrading, and maintaining the data systems.

**Internal EPA Review
Proposed Restructuring of
Information Responsibilities**

For example, EPA charged a task force of senior managers with redesigning the agency's internal management structure to better meet its new information demands. In carrying out its charge, the task force consulted with EPA employees, external stakeholders, and the states. The report of the task force recommended establishing a single program manager for information management and policy combined with strengthening information resources management and technology functions. In response to the recommendations, the agency has planned the establishment of an Information Office that would bring together information management functions previously housed separately. The new office is to contain a new Quality and Information Council with a role that includes the provision of agencywide strategic direction and advice on quality and data collection. An Information Quality Staff is to support the Council.

**DOT Legislation Required
External Panel Review of
Data Collection Capabilities**

Legislation establishing the Bureau of Transportation Statistics (BTS) called for the National Academy of Sciences to conduct an external review of the adequacy of data collection procedures and capabilities of the Department of Transportation (DOT). A panel of experts was subsequently appointed to examine the functions that BTS could or did perform and the resources and capabilities it had to carry out those functions. The study report noted a number of areas for improvement, including the need to build and maintain a strong statistical and technical staff, which is being implemented by DOT.

**Facilitate Agencywide
Coordination and
Cooperation**

Coordination of data quality efforts across systems or offices can be a key issue in agencies. Reporting on annual performance goals may require integrating data from different systems, which requires coordination across organizational units.

The agencies we contacted have numerous data systems that collect the data used for agency performance measures. We found that, often, these systems were initially constructed to meet the management needs of specific programs, sometimes in another organizational unit. Consequently, they may collect different data elements or use different data definitions and standards, even for the same data element. The involvement of higher level administrators may be needed to facilitate the necessary coordination among semi-independent organizational units and to obtain agreement on the division of responsibilities. The agencies we reviewed had developed a variety of mechanisms to facilitate the

coordination and cooperation needed for good quality data in these circumstances.

VHA’s Data Quality Council to Ensure Coordination

For example, the Veterans Health Administration has established a Data Quality Council, chaired by the Deputy Under Secretary for Health, with a mandate to ensure open discussion and greater collaboration in the area of data quality policy development and implementation. The Council is to include representatives from headquarters as well as regional and field offices. Among its specific responsibilities is ensuring the coordination and communication of major national data quality issues and initiatives.

Several ED Groups to Coordinate Agencywide Implementation of Data Standards

The Department of Education (ED) is initiating several activities to foster agencywide implementation of data standards. Several groups contribute to the coordination of this effort, including a strategic planning team; a panel to review the individual performance plans submitted by each ED office, including a review of data sources and quality; and a work group to develop data quality standards.

OPM Uses Work Group to Coordinate Performance Measure Development

Coordinating the perspectives of multiple organizational stakeholders may enhance the validity of data elements chosen to provide a performance measure. For example, the Office of Personnel Management (OPM) used a work group involving several organizational units in developing and approving survey items for assessing OPM’s performance on federal personnel policy issues.

Assign Clear Responsibilities for Various Aspects of the Data

Many people “touch” performance data, including suppliers and creators of the data, those who store and process data, and those who use it. Data are more likely to be of high quality when it is clear who is responsible for each step in data creation and maintenance, from the initial specification and definition of data elements to correctly entering data about clients; providing training for and supervision of those who enter data; transferring data from initial to final formats; and appropriately analyzing and reporting the performance measures.

A primary responsibility for the quality of a program’s data rests with the manager of that program. Often, performance information is directly collected by the operating components and may be used for managing those programs. Because these data are also used for decision-making by other levels of management and for reporting to Congress, managers’ direct responsibility for data quality has broader implications. Several agencies are explicitly holding immediate program managers and their divisional administrators accountable for the quality of data from their programs.

ED Is Requiring Managers to Certify That Their Data Meet the Standards

For example, the Department of Education is planning to hold managers accountable by requiring them to attest to the quality of the program data used for performance measures. ED has developed detailed data quality standards and procedures for implementing this requirement. ED's evaluation office will also provide support services, such as training in the application of data standards for performance measurement. If they cannot certify that the data for a performance measure meet the standards, the managers are to provide plans for bringing the data up to standard.

EPA Assigns Responsibility to Sponsors and Producers of Data

EPA provides an example of agency efforts to clearly assign responsibilities for various aspects of data quality. Headquarters' sponsors of data in EPA's Comprehensive Environmental Response, Compensation, and Liability Information System database are responsible for identifying and defining needed data elements, and the regional manager who produces the data is responsible for reviewing, verifying, and validating the data for this system. An Information Management/Program Measurement Center under EPA's Office of Emergency and Remedial Response is assigned a variety of responsibilities for the completeness, accuracy, integrity, and accessibility of data.

Adopt Mechanisms That Encourage Objectivity and Independence in Collecting and Managing Data

An organizational capacity for objectivity and independence in key data collection, management, and assessment processes can help create a climate that fosters data quality. Fostering objectivity and independence as a protection against bias is a major principle of several disciplines, including auditing, scientific research, and program evaluation.

OPM's RIS Operates Independently From Program Offices

We found instances of a deliberate management strategy to introduce mechanisms for fostering independence in data collection and management. At the Office of Personnel Management, the Retirement and Insurance Service's (RIS) Management Information Branch operates independently from the relevant program offices and reports directly to the RIS Associate Director through the Assistant Director for Systems, Finance, and Administration. OPM's fiscal year 2000 performance plan notes that this arrangement is part of its strategy "to ensure the integrity of the performance indicators" derived from its comprehensive management information system, which is used to monitor and report output (business process) measures.

ED Is Planning Several Mechanisms for the Independent Review of Data

At the Department of Education, in addition to holding program managers responsible for their program’s data, several mechanisms are being planned to ensure independent review of the data submitted, including involvement and review by staff of the National Center for Educational Statistics, Inspector General’s Office, and Planning and Evaluation Service.

Provide Responsible Staff With Training and Guidance for Needed Skills and Knowledge

Both external sources and our prior publications concerning results management have emphasized the importance of training managers about measurement issues so they can implement performance management.² Understandably managers whose prior job roles emphasized other tasks, such as awarding and managing portfolios of grants or contracts, may lack skills in using performance data—that is, managing for results. Without assistance, management and staff may not understand how measurement is best implemented or how to correctly interpret performance information, especially how to acknowledge its limitations.

We have previously noted that “. . . staff at all levels of an agency must be skilled in . . . performance measurement and the use of performance information in decision-making. Training has proven to be an important tool for agencies that want to change their cultures.”³

ED Plans to Train Managers

The Department of Education’s Inspector General reported on the status of ED’s implementation of the Results Act before March 1998. The Inspector General found that “not having a sufficient number of staff qualified in information processing, evaluation and reporting,” and “the difficulty of analyzing and interpreting performance measurement data” were two of the three most frequently identified barriers to successful implementation of the Results Act identified by 27 key staff they interviewed. ED plans to provide several types of training opportunities for its managers, including conducting workshops provided by a training contractor, one-on-one “coaching” with managers of the largest programs, and having evaluators review managers’ self-ratings of data quality to ask questions about weaknesses they may have overlooked.

²Kathryn E. Newcome, “Comments on the Future for Performance-Based Management in the U.S. Federal Government” in Federal Committee on Statistical Methodology, Office of Management and Budget, Statistical Policy Working Papers (#28: Seminar on Interagency Coordination and Cooperation), pp. 148-49, and Report of the Auditor General of Canada, “Moving Toward Managing for Results” (Oct. 1997), pp. 11-26.

³Executive Guide: Effectively Implementing the Government Performance and Results Act (GAO/GGD-96-118, June 1996), p. 42.

The VA Inspector General Has Issued a Program Manager's Guide on Auditable Performance Measures

Some agencies are already developing training and guidance for managers to build the needed knowledge and skills. For example, the VA Office of the Inspector General has issued a program manager's guide on the auditor's approach to auditable performance measures. By highlighting the requirements for data collection and the common problems found in its audits, the Inspector General intends to contribute to program managers' awareness and better enable them to plan for their data collection and performance reporting activities.

An Example of Agencywide Capacity-building at ED

The approaches we described in this section can be part of an integrated strategy for fostering the agency's commitment to enhanced data quality. As discussed, the Department of Education is embarking on a major effort to make high-quality performance data an agencywide priority. A summary of its plans illustrates an integrated strategy that incorporates multiple approaches.

Senior ED managers have made obtaining valid and verifiable data a departmental priority. This is documented by the inclusion of the following indicator in ED's fiscal year 2000 performance plan: "By 2000, all ED program managers will assert that the data used for their program's performance measurement are reliable, valid, and timely, or will have plans for improvement.

Several groups provide related coordination. These include a strategic planning team; a panel to review all performance plans, including a review of data sources and quality; and a work group to develop data quality standards.

In carrying out this priority, the Department of Education

- has developed an explicit set of "Standards for Evaluating the Quality of Performance Indicators/Measures," which includes definitions, examples of meeting or failing to meet each standard, and possible methods for data checking, for each of six standards;
- is requiring program managers to certify the data quality for each performance indicator, using a standard rating system, or to provide plans for bringing data quality up to the standards;
- has developed an explicit plan for implementing the data standards, which documents the detailed steps to be followed;
- is providing several types of training for program managers on the data standards and their implementation;
- is using independent oversight by the evaluation office and the Inspector General to provide concurrence with program managers' assessment of the

quality of their data and to issue a Program Data Quality Report Card summarizing data status;

- is developing new integrated data systems for elementary and secondary data to coordinate data definitions, collection, and reporting among ED's programs and state and local education agencies; and
- is communicating with Congress on statutory changes needed to support the Department's reporting for the Results Act; for example, making recommendations for the reauthorization of the Elementary and Secondary Education Act to limit data required from states to the data elements essential for Results Act reporting.

Assessing the Quality of Existing Data

One strategy for verifying and validating proposed performance data is to assess the quality of current data to identify problems that may affect their use. Assessments might target specific measures in the performance plan or more broadly assess major data systems and their problems. Examples of both broad and narrow data assessments are presented below. If an assessment shows adequate data quality, the agency will need to reassess performance data periodically to verify that data quality is maintained. Such quality assessments can examine each measure along the relevant dimensions of data quality, such as those described in appendix I, for example. In assessing data quality, the agency may first need to establish the appropriate quality dimensions for its data because these may differ for various types of programs and data sources.

Several approaches for assessing existing performance information are shown in figure III.1 and discussed below. These are intended as an initial “menu of approaches,” not a checklist of requirements. Each agency needs to choose approaches that fit the intended uses of its performance information, nature of its data and data systems, resources available for assessing the data, and initial diagnosis of the extent of problems in current data systems.

Figure III.1: Approaches to Assessing the Quality of Existing Data

- Build data quality assessment into normal work processes, including ongoing reviews or inspections.
- Use software checks and edits of data on computer systems and review their implementation.
- Use feedback from data users and other stakeholders.
- Compare with other sources of similar data or program evaluations.
- Obtain verification by independent parties, including the Office of the Inspector General.

Build Data Quality Assessment Into Normal Work Processes, Including Ongoing Reviews or Inspections

For some types of performance measures, quality assurance procedures can be built into the agency’s normal workflow and managerial oversight. These may be appropriate, for example, when the performance data are derived from information systems used to manage the workflow, such as benefits or claims determination. This approach is consistent with advice from the business world to design quality into data systems for managing business processes by using methods that make data error detection and correction a normal part of agency operations.

OPM's RIS Routinely Verifies the Accuracy of Its Calculations and Provides Feedback to Managers

For example, the Office of Personnel Management's (OPM) Retirement and Insurance Service (RIS) routinely verifies the accuracy of federal retiree benefit calculations. Several times each year, its Quality Assurance Division reviews a statistically representative sample of completed calculations and draws conclusions as to whether each claim was merited or not. Division staff provide feedback from this verification to help managers maintain quality control over the accuracy of benefits awarded. OPM staff stated that they now have a 95-percent accuracy rate from this process. The aggregate data from these accuracy checks are also now used as a performance measure to meet the requirements of the Results Act.

VBA Annually Reviews Cases Selected Randomly From Regional Offices

A major role of the Department of Veterans Affairs' (VA) Compensation and Pension programs is to process disability claims from veterans. This program has built accuracy checks into its normal work processes since 1992. Under a new quality review system, the Systematic Technical Accuracy Review (STAR), VBA headquarters will annually review about 7,400 cases selected from regional offices. In addition, VBA will require each regional office to review samples of its own work products using STAR procedures. The purpose of STAR is "to improve the accuracy of compensation and pension claims by providing current and diagnostic information about the accuracy of work being produced at the field stations."¹Data from the STAR system will also be used for the program's performance reporting.

EPA's Science Advisory Board Conducts Peer Reviews

The Environmental Protection Agency's (EPA) Science Advisory Board was established to provide independent scientific and engineering advice to the EPA Administrator on the technical basis for EPA regulations. Its members include scientists, engineers, and economists from academia, industry, and the environmental community. The board conducts scientific peer reviews to assess the technical merit of agency positions. These reviews include whether data are of sufficient quality to support environmental measures and whether proposed measurement models and methods are appropriate.

¹Department of Veterans Affairs, Fiscal Year 2000 Budget Submission, Departmental Performance Plan, p. 49.

Use Software Checks and Edits of Data on Computer Systems and Review Their Implementation

Use of electronic data management and processing systems normally includes a variety of automated checks on data values and integrity. These can include built-in “range checks,” which notify the data manager if an entered value falls outside of the expected range for that data element; consistency checks among several data elements that should be consistent, such as age and current enrollment in school; procedures for ensuring the integrity of data files and their links to other files; and overall system controls for data security and integrity.

Assessing the quality of existing data in such computerized systems involves reviewing the implementation of these procedures. Such a review would start with the detailed documentation for each data system to assess the completeness of the intended software checks and edits. The actual procedures would then be reviewed to ensure that they are being carried out consistently. Finally, the results from the data checks would be examined. The assessment would include such things as the percent of the data initially entered that fall outside the range and consistency checks and the implementation of procedures for correcting the data.

Our publication, Assessing the Reliability of Computer-Processed Data (GAO/OP-8.1.3, Sept. 1990), discusses the issues concerned with this approach, and provides several checklists of useful types of data tests and computer system controls in its appendixes I and II.

Contractors Must Apply ED’s Standards on Data Preparation and Reporting

The Department of Education’s (ED) publication on Standards for Education Data Collection and Reporting contains a major section on data preparation and reporting, with nine detailed sets of standards for these processes, including designing data processing systems, testing data processing systems, and documenting data processing activities. ED staff indicated that the contractors who conduct the projects used to provide data for performance measures must apply these standards.

Use Feedback From Data Users and Other Stakeholders

Current users of data systems and their results may have valuable experience with the strengths and weaknesses of existing data and can provide insights into the data’s credibility with external audiences. These users and stakeholders can include agency staff members in program or statistical offices; providers of data, such as state agencies or local grantees; academics or “think tank” staff who use the data for policy analysis; and industry representatives who base plans or decisions on comparative or trend statistics from the data.

EPA Has Several Methods for Obtaining Data User Feedback

For example, the Environmental Protection Agency seeks stakeholder feedback on the quality and usefulness of its performance data in several ways: customer consultations, posting feedback forms on its Internet site, and sending data to users and providers for verification.

To obtain “customer” feedback, the Center for Environmental Information and Statistics conducted meetings with national, regional, state, and local environmental data users to ask what information they need and how they would like to access it. In addition, the participants expressed concerns with the accuracy of data entry, transmittal, and agency reporting.

EPA posts a wide variety of environmental information on its Internet Web site. In particular, the site’s Envirofacts Warehouse provides a single point of access to environmental data maintained by EPA. The site and each data source link to a feedback form that invites questions or comments about Envirofacts databases.

EPA also verifies some data by sending it back to its originators for comment. The agency’s pilot Sector Facility Indexing Project includes information on inspections of regulated facilities and noncompliance with regulations. As part of its process for verifying the data, EPA sent each facility a copy of its compliance and enforcement data for review and comment to make sure mistakes were caught before the information was released.

VHA Convened a Data Quality Summit to Elicit Feedback From Stakeholders

Using stakeholders to provide feedback on data collection and management problems, the Veterans Health Administration (VHA) convened a 3-day Data Quality Summit in December 1998 to bring together staff from its network of veterans hospitals, information systems, and central office staff. Prior to the summit, participants were asked to prepare papers on data quality issues that they believed affected the organization. Examples of issues identified by participants included coding problems, data definitions, and data correction and consistency. The Data Quality Summit obtained input on potential solutions that would meet the needs of the multiple users of the data systems. Follow-up work groups are to develop action plans for needed improvements.

NSF Uses a Contractor to Check Validity of Data Reported by Projects

For one of the National Science Foundation’s (NSF) science education programs, a contractor was used to obtain feedback on the validity of data reported annually by each project. To confirm that these data elements incorporated the intended meaning, a contractor conducted an informal telephone survey of 15 projects, asking project evaluators about their understanding of the questions used in reporting the data items. The

contractor also collected more detailed information about the procedures used in collecting the data locally and identified problems that projects were having with individual items. The contractor reported that respondents generally understood the data definitions, concepts, and time frames that had been established to govern responses to individual items.

Compare With Other Sources of Similar Data or Program Evaluations

A fourth approach to assessing the quality of existing data is to compare the performance information with data from other sources. Comparisons can serve several different purposes. One purpose is to assess the validity of the performance measure by comparing the data elements or source to be used with related data elements from another source. Another purpose is to test the accuracy of data from an ongoing system with data from a more rigorously collected source that may be available only periodically, such as a one-time evaluation.

OPM's RIS Compares Management Information System Data With Periodic "Physical Inventories" of Paper Documents

One example is provided by the Office of Personnel Management's Retirement and Insurance Service, which processes about 4 million paper items each year to manage the federal retirement system. To cross-check the accuracy of the performance statistics in their management information system, the central office staff reports that they periodically request a "physical inventory" of pending work in each local office at the end of a week. They compare actual counts of hard-copy documents on hand at that point with that office's statistics generated by the management information system for that week. If there are discrepancies, the central staff works with local managers to avoid duplicate counting and other errors.

VHA Compared Data From Program Offices With Centrally Aggregated Data

Useful comparison among data sources can include analysts' judgments. Staff at the Veterans Health Administration reported that they compare data from program offices with more aggregated data from their central systems for assessing health care "capacity" indicators. If the comparison reveals inconsistencies in these sources, they reconcile differences by contacting the relevant program managers to learn reasons for the differences and to reach consensus on the most accurate numbers for the intended presentation.

Obtain Verification by Independent Parties, Including the Office of the Inspector General

For data elements drawn from data systems in regular use, a key assessment step could be the verification of the accuracy of results by an external, independent examiner, such as a professional body or the agency's Inspector General. A reported result can be checked for completeness, consistency, and accuracy by tracing the data, or a representative sample of data, back to their original source. Verification can involve analyzing whether the end data match the initial data source. It

can also involve assessing whether data collection and transformation procedures are fully documented and followed consistently.

ED's Inspector General Traced Data Flow and Control in Several State Educational Agencies

For example, several measures for the Department of Education's elementary and secondary education assistance programs will be provided by state educational agencies (SEAs). To assess the accuracy, completeness, methods of calculation, and presentation of targeted educational data elements for example states, ED's Inspector General conducted field assessments in four SEAs in early 1999. The Inspector General's staff conducted interviews with state officials and attempted to trace the data flow and data control processes in place in each state.

This exploratory work toward data verification is intended to identify processes used by SEAs to accumulate and report performance data to ED, to identify limitations in the data submitted, and to describe any barriers to improving data quality. This assessment also provides background for a major redevelopment of joint data collection efforts between ED and its state and local partners.

VA's Inspector General Is Assessing Critical Data Elements

The Inspector General for the Department of Veterans Affairs has focused on audits of key performance measures. With input from management, the Inspector General identified a subset of 11 performance measures considered most critical for measuring the agency's performance. The initial audit focused on data for three measures relating to the timeliness achieved by the Veterans Benefits Administration in processing claims from veterans for disability compensation and pension benefits. The audit assessed the data for validity, reliability, and integrity (the extent to which the data could not be "gamed" or manipulated), in accordance with guidance contained in our report, [Assessing the Reliability of Computer Processed Data](#) (GAO/OP-8.1.3, Sept. 1990).

The Inspector General compared source documents with information on automated systems for three random samples of claims completed in fiscal year 1997. The audit found that "more than 30 percent of the records in each of the three samples contained inaccurate or misleading data." VBA administrators have cited the findings as an impetus for rigorous data improvement efforts.

OPM Uses External Certification of Health Care Data

Use of data that are "certified" by an external, professional body is another means for independent verification. For example, the Office of Personnel Management, which administers the federal employees' health insurance program, works closely with the professional organization for improving quality in managed health care, the National Committee for Quality

Assurance (NCQA). OPM requires its health insurance carriers to submit scores for the Health Plan Employer Data and Information Set (HEDIS), which is managed by NCQA. HEDIS is a set of standardized health care quality measures used to compare managed care health plans.

To ensure that HEDIS quality specifications are met, NCQA has developed a data auditing procedure using licensed organizations and certified auditors for assessing carriers' nonfinancial data elements. The audit includes verifying a sample of HEDIS measures to confirm that HEDIS results are based on accurate source information. The process results in a certification rating of "Report," "Not report," or "Not applicable" for each measure reviewed.

Consequences From Assessing the Quality of Existing Data

As a result of using the approaches outlined above, or from other data assessment procedures, the agency will be able to identify data of adequate quality for some measures, as well as gaps and limitations in some data elements planned for use as performance measures. For some of the limitations, the approaches identified in appendix IV can be undertaken to provide more credible data. In other circumstances, the agency may decide that it needs to substantially change its data acquisition process or create a new data system to "build quality" into performance data, which is addressed in appendix V.

Responding to Data Limitations

Agencies have undertaken a variety of approaches to assessing the quality of existing data, as discussed in appendix III. Assessments of data quality do not lead to improved data for accountability and program management unless steps are taken to respond to the data limitations that are identified. Guidance for assessing agencies' performance plans calls for them to identify significant data limitations and to discuss the steps being taken or proposed to address those limitations.¹ In the report summarizing observations on 1999 agency performance plans, we found that "in general, agencies' annual performance plans did not include discussions of known data limitations and strategies to address them."² Our assessment of the fiscal year 2000 plans found that agencies generally do not identify actions they are taking to compensate for the lack of quality data, nor do they discuss implications for decision-making.³

Improving future performance information, as outlined in appendix V, is one important response to findings concerning data limitations. Appropriate agency responses to directly address the data limitations, shown in figure IV.1, are discussed below.

Figure IV.1: Approaches for Responding to Data Limitations

- Report data limitations and their implications for assessing performance.
- Adjust or supplement problematic data.
- Use multiple data sources, with offsetting strengths and limitations.
- Improve the measure by using another source or new methods of measurement.

Report Data Limitations and Their Implications for Assessing Performance

Making stakeholders aware of significant data limitations allows them to judge the data's credibility for their intended use and to use the data in appropriate ways. All data have limitations that may hinder their use for certain purposes but still allow other uses. Stakeholders may not have enough familiarity with the data to recognize the significance of their shortcomings. Therefore, appropriate use of performance data may be fostered by clearly communicating how and to what extent data limitations

¹The Results Act: An Evaluator's Guide to Assessing Agency Plans (GAO/GGD-10.1.20, Apr. 1998), p. 45.

²GAO/GGD/AIMD-98-228, Sept. 8, 1998.

³GAO/GGD/AIMD-99-215, July 20, 1999.

impact on assessments of performance. Communicating the implications of data limitations can involve specifically identifying appropriate and inappropriate interpretations of the data.

DOT's BTS Reports on Data Sources, Gaps, and Weakness

In response to a legislative requirement that the Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) identify information needs on an ongoing basis, the Bureau published a report that identified initial gaps in transportation statistics and proposed strategic responses. The report identifies data gaps and weaknesses in a variety of areas, such as transportation safety; energy use; and the flow of people, goods, and vehicles. For example, the report notes that transportation injuries are underreported and that there are inconsistencies in how injuries are reported, complicating the assessment of transportation safety.

ED Provides a Section on Data Limitations for Each Indicator in Its Annual Plan

The Department of Education (ED) includes a section on "Limitations of the Data" when presenting each indicator in its performance plan. (See fig. I.1 in app. I for ED's presentation format.) For some objectives, ED also discusses the reasons for and implications of these limitations as they affect the verification and validation of the measure. In addition, as part of their review and certification that data for performance measures meet ED's data quality standards, program managers are to identify any standards that are not met and steps for correcting these limitations.

DOT's Performance Plan Explains How Limited Data May Still Be Useful

An appendix in DOT's performance plan for fiscal year 2000 contains a section on limitations in the data sources for each performance measure in the plan. The discussions of limitations for some performance measures also include the implications of the limitations for performance measurement. For example, the plan notes that because of the judgment involved in assessing whether mariners are in distress, the reported rate may overestimate the number of lives saved. However, the plan argues that the reporting from year to year is likely to be consistent, providing a reasonable estimate of changes over time.

ED Describes Challenges Involved in Obtaining High-Quality Performance Data

In addition to describing some of its data limitations, ED's performance plan provided a context for its efforts to address limitations by describing the challenges that they faced. A detailed section on "measurement challenges" describes data limitations derived from the decentralized system of elementary and secondary education, in which many national goals and objectives are under limited federal control. Further, it discusses the need to measure programs with overlapping goals but disjointed information systems as well as identify knowledge gaps where the Department is attempting to "measure the hard-to-measure."

Adjust or Supplement Problematic Data

Sometimes, data limitations can be overcome by conducting accepted statistical adjustments, such as statistical modeling or estimating values for missing data elements. However, statistical adjustments are sometimes controversial and can be hard for nonspecialists to understand. Appropriate use depends on a number of assumptions underlying each adjustment procedure, whose application requires considerable specialized expertise.

DOT Uses Statistical Adjustments to Compensate for Missing Data

One common data limitation is the inability to get information on all cases of interest. DOT reports the use of statistical adjustments to compensate for this problem. Blood alcohol consumption test results are not available for all drivers and nonoccupants involved in fatal crashes. Using important crash characteristics, such as crash vehicle and person factors, the DOT's National Highway Traffic Safety Administration (NHTSA) seeks to avoid an undercounting of these fatalities by employing a statistical model to adjust for missing data. Without this correction, the percentage of alcohol-related highway fatalities would appear to be lower than they actually are.

ED Analyzes Nonresponses and Statistically Adjusts the Reported Data

The Schools and Staffing Survey, used by ED for several performance measures, collects data from a variety of educational staff, such as teachers, administrators, and librarians and for public and private schools. Even after rigorous survey administration procedures and telephone follow-up, response rates differ among these components. To reduce bias in reported results, the National Center for Educational Statistics conducts analysis of the sources of nonresponse, then uses statistical procedures to adjust the data reported.

Use Multiple Data Sources With Offsetting Strengths and Limitations

Comparing information derived from data sources with different strengths and weaknesses adds confidence to judgments about performance. Agencies may have access to two or more data sources that can provide information on a given area of performance. Although each data source may have serious limitations, confidence in results may be increased when each source provides the same overall picture of performance. Combining data sources may also provide a more complete picture of performance than can be obtained from a single source.

OPM Is Identifying Consistencies and Discussing Differences in the Results of Three Surveys

The Office of Personnel Management is comparing results from three different surveys to identify consistencies and to stimulate discussion of reasons for any differences. The surveys examine federal employee and human resource personnel satisfaction or perceptions with regard to human resource operations and OPM initiatives. They expect to report on these analyses in next year's performance plan.

OPM Customer Satisfaction
Data Validated by Web
Master

OPM's Employment Service operates an automated Employment Information Service, containing postings of federal job openings. The system automatically collects several kinds of data, including use statistics and customer satisfaction feedback. The results concerning customer satisfaction are validated qualitatively by the system Web master from complaints received and any technical problems identified with regard to recent system enhancements.

Improve the Measure
by Using Another
Source or New
Methods of
Measurement

Some data limitations can be addressed by replacing the data source. In some cases, improving data collection and management procedures, as described in appendix V, may correct the problem. Comparing data with its original source and correcting the errors in existing data may also be possible, for example, if the limitations occur because of inaccurate data coding and entry. However, fixing existing data can be expensive, and unless stakeholders require the historical data as a baseline, the resources may be better used to find new information or new methods of measurement.

NHTSA Uses Data From
Private Industry

The National Highway Traffic Safety Administration uses private industry data on vehicle registration rather than federal estimates, believing that the former more closely reflect the actual mix of vehicles on highways. Federal statistics are obtained from state information systems, which may overcount certain vehicles if they have been transferred from one state to another and show up in both states' files.

VBA Has Changed Its
Method for Estimating the
Accuracy Rate for Claims
Processing

The Veterans Benefits Administration (VBA) has recently changed its method for estimating the accuracy rate for the processing of veterans' compensation and pension claims. This rate is one of its performance measures. VBA's system for measuring accuracy had indicated an estimated 95-percent accuracy rate for the claims processing activity. However, questions arose because the processing of veterans' appeals of these initial decisions reversed about 19 percent of the appealed decisions and remanded about 47 percent back for further development and reconsideration.

The Systematic Technical Accuracy Review (STAR) was implemented to improve the accuracy of the work of compensation and benefits officers and to provide information for measuring annual performance goals concerning accuracy. Pilot tests of the new STAR system found only a 64-percent accuracy rate in claims processing decisions. Compared to the earlier system, the STAR system focuses more on decisions that are likely to contain processing errors and uses a stricter standard for computing accuracy rates.

Building Quality Into the Development of Performance Data

Improving data quality by detecting and correcting errors with existing data will not necessarily prevent future errors. Assessments of existing data elements or systems to be used for performance measures may reveal that improvements are needed in current data systems or that new systems are needed. Agency performance plans are expected to indicate any changes or improvements being made to modify, improve, or expand the capability of existing data systems or processes, according to the Office of Management and Budget's (OMB) Circular A-11 guidance for performance plans.

Reporting data design and collection procedures may be particularly useful when data are collected episodically, rather than on an ongoing basis. In these circumstances, it may not be feasible to verify the data by comparing them to an original source or alternative data sources, such as in a nonrecurring survey.

Figure V.1 lists approaches that agencies can take to build quality into their performance data.

Figure V.1: Approaches to Building Data Quality Into the Development of Performance Data

- Use prior research or analysis to identify data elements that adequately represent the performance to be measured.
- Gain agreement among internal and/or external stakeholders about a set of measures that are valid for their intended use.
- Plan, document, and implement the details of the data collection and reporting systems.
- Provide training and quality control supervision for all staff who collect and enter data, especially at local levels.
- Provide feedback to data collectors on types of errors found by data checks.
- Use analytic methods and transformations appropriate for the data type and measure being reported.

Use Prior Research or Analysis to Identify Data Elements That Adequately Represent the Performance to Be Measured

Several agencies use findings from basic research to assess the validity of potential data elements for measuring intended performance. Research may illuminate the relationships between the agency's strategies and outcomes in the content area of the performance measure. Or, appropriate measuring tools and data collection procedures may be drawn from this literature or adapted to become more compatible with the agency's needs.

EPA Uses Research on Health Risk in Drinking Water

For example, the Environmental Protection Agency's (EPA) Safe Drinking Water Program uses research on the health risks associated with specific levels of exposure to set standards for maximum contaminant levels. The agency measures its annual progress in ensuring that Americans will have clean and safe drinking water by estimating and reporting the percentage of the population served by water systems that meet all health-based standards.

OPM Used Research on Customer Satisfaction

The Office of Personnel Management (OPM) uses several sample surveys to assess federal agency human resources staff satisfaction, for example, with OPM technical assistance and guidance materials. To develop valid items for its survey instruments, OPM's Personnel Resources and Development Center reviewed extensive research on "customer satisfaction" in the fields of organizational psychology, management, and marketing. From this literature, OPM identified nine underlying service dimensions of customer satisfaction, including the courtesy, knowledge, and timeliness of the service staff as well as the extent of choice and quality for the specific service. OPM developed a set of survey scales with 30 core items for these nine dimensions, along with four general items about overall quality and satisfaction. The core items were pretested with staff in three agencies before the measures were included in OPM's customer satisfaction surveys, used to provide measures in OPM's performance plan.

Gain Agreement Among Internal and External Stakeholders About a Set of Measures That Are Valid for Their Intended Use

Selecting or developing valid data elements can also be enhanced by involving others who collect or use the resulting data (stakeholders). This step is particularly useful when staff outside the agency will be the primary data collectors, such as staff in state or local agencies or grantees. Such consultation helps to establish consensus on the data elements that are valid measures of the underlying concept and that take into account the varied local circumstances and resource availability affecting the consistency of data collection.

An ED Panel Is Developing an Integrated System for State Reporting

For example, the Department of Education (ED) is facilitating the development of an Integrated Performance and Benchmarking System (IPBS) for its elementary and secondary education programs. Currently, most ED programs have separate reporting systems, with considerable overlap in the types of information collected, but not always common definitions of key terms, such as “student.” Most states also collect similar types of data, but often not in ways that allow them to compare with other states. The IPBS initiative will seek agreement among states and ED program managers for a common core of data elements. It is currently in an exploratory phase, with representatives from two states cochairing a panel in conjunction with staff from ED to develop a system plan. Full national implementation is intended by 2004. ED also expects to award financial grants to states for implementing the needed improvements.

Further, such consultation may be desirable even within an agency, to avoid overemphasis on any single measure. Collecting data on only a limited aspect of total performance may encourage management and staff to look for ways to make performance appear better than it actually is. Obtaining within-agency agreement on a more balanced set of performance measures may help to minimize distortions that can result from overemphasis on a single measure.

VBA Is Developing an Expanded, More Balanced Set of Performance Measures

For example, Veterans Benefits Administration (VBA) officials told us that improved data quality was an anticipated benefit of their adopting a “balanced scorecard” approach to performance measurement. For this approach, VBA staff are developing an array of measures that capture the various elements of VBA’s strategic vision, including measures of timeliness, accuracy, unit cost, employee development, and customer satisfaction. The new set of measures expands VBA’s previous emphasis on timeliness and productivity. Although improved data quality is not the primary purpose for adopting a more balanced set of measures, the officials we talked to believed that this would be one benefit of the approach.

Plan, Document, and Implement the Details of the Data Collection and Reporting Systems

Agencies find that developing new or revised data systems involves a number of aspects that need to be carefully planned and carried out for the resulting data to be valid and verifiable. These aspects can include the exact specifications of the data elements to be collected, the population or sample of entities from which to collect data in each location, the detailed steps for data collection and manipulation in each location, training for local data collectors, oversight procedures for local supervision of data collection, and quality standards to be employed in that oversight. After these efforts, the subsequent reporting of validation and verification

methods in the performance plan could focus on the methods employed to build in data quality, for example, by documenting the steps undertaken in developing and implementing the data collection system.

ED's Even Start Family Literacy Program Illustrates Planning for Data Definitions and Data Collection

To obtain consistent data from different locales, detailed plans for the data definitions and data collection procedures are needed. This planning may involve local partners, if they will be responsible for data collection. For example, ED's Even Start Family Literacy Program provides a multicomponent program for low-income families in more than 600 locations across the nation. ED staff report they have involved the grantees in developing data definitions and changes in the data collection procedures as they have evolved since 1989. ED uses a contractor to work with the local projects in developing its reporting system, which has five data reporting forms for various population groups in the program. The data collection system is documented in a detailed user's manual, which contains an explanation of every question in every form, as well as instructions for using the automated data entry system. The contractor maintains a toll-free telephone line for answering questions about the data forms and communicates immediately with a grantee if its data submission appears to contain errors.

Plans for data processing at a central level also need to be developed and documented to ensure consistency among multiple staff and over time, as turnover occurs among staff. These plans include how the data will be transferred from the individual collection sites, how it will be stored and processed, and how it will be aggregated into the needed performance measures. These developmental steps involve the technical staff and data processing specialists from the several organizational levels that will collect and manage the data. Implementing the plans will also involve using the software checks and edits for data on computer systems that are discussed in appendix III.

EPA's Toxic Release Inventory Illustrates Planning and Control of Data Processing at the Central Level

For example, EPA's Toxics Release Inventory is a database containing industry-reported data or estimates about releases of listed chemicals that exceed certain amounts. EPA's Center for Environmental Information and Statistics indicates that every facility uses the same forms for reports and that input forms are checked centrally for completeness, valid formats, chemical identification numbers, and internal consistency. The agency runs computer checks against the reported data. When potential errors are identified, facilities are notified to allow for correction.

ED Developed Quality Standards for Performance Measures Using Relevant Expertise and Professional Standards

The development of new or revised data systems can be aided by utilizing relevant expertise and professional standards to advise on both the content of the measures and the technical aspects of information systems. For example, ED developed a brief set of draft “Standards for Evaluating the Quality of Performance Indicator Measures” that all the Department’s programs will be required to follow when reporting their performance data. To develop these standards, ED drew on internal expertise from several disciplines, including both educational statistics and auditing; used a contractor to collect examples of quality standards; and had draft standards reviewed intensively by the Department’s Evaluation Review Panel, a group of external evaluation experts from academia and state agencies.

EPA Environmental Data Collection Must Demonstrate Conformity With ANSI Standards

EPA requires all its environmental programs to be supported by quality systems that comply fully with standards for environmental data collection developed by the American Society for Quality and authorized by the American National Standards Institute (ANSI).¹ EPA’s policy requires the development of Quality Management Plans for programs and Project Assurance Project Plans for individual projects, as recommended in the standards.

ED’s NCES Has Detailed Survey Standards and Specifications

The ED’s National Center for Education Statistics (NCES) has detailed standards and specifications for designing, conducting, and analyzing educational surveys, including those collecting data used for performance measures.² These standards are built into new contracts for data collection, and quality control procedures are monitored by each contract’s technical officer, then documented in technical reports for each survey. Each project also has a technical review panel, which reviews the details of survey design and quality control during data collection.

NSF Is Using an Electronic Web-Based Process to Obtain Final Reports From Grantees

Some agencies are trying to minimize data entry and transmittal errors by using or planning for electronic data systems, rather than using paper-based data collection forms, for initial data entry and transmittal to a central location. For example, the National Science Foundation (NSF) is implementing an electronic, Web-based process for the submission of final reports from its research grants. The on-line report format includes a number of features to ensure appropriate data entry, including hypertext

¹American Society for Quality Control, American National Standard: Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs (ANSI/ASQC E4-1994, Jan. 3, 1995).

²National Center for Education Statistics, NCES Statistical Standards (Washington, D.C.: Department of Education (NCES 92-021r)), 1992; Westat, Inc., and NCES, SEDCAR: Standards for Education Data Collection and Reporting (Washington, DC: Department of Education (NCES 92-022r)), 1991.

explanations and definitions as to what is needed in each entry, automatic range checks to flag improper entries, and immediate feedback to the grantee regarding invalid entries.

The electronically submitted reports are then reviewed by NSF's content area project officers, who check whether entries are reasonable and consistent with other information about that project. NSF staff believe that this electronic report submission leads to quicker turnaround time for data submission; simplified and more accurate data entry; data more relevant for program management; and therefore, more use of the data by program managers.

Provide Training and Quality Control Supervision for All Staff Who Collect and Enter Data, Especially at Local Levels

The quality of any information system, and of the performance measures derived from it, depends on the quality of the data being entered. To obtain the necessary consistency and accuracy in data collection and entry, several agencies are providing training and supervision of data collectors as a part of their data quality procedures. Such training helps to ensure that those collecting and entering the data have a common understanding of the meaning of each data element, the conventions for using each categorization or coding rule, the frequency with which data are to be entered, and so on. If data collection will use electronic forms, hands-on experience with sample cases to code and enter during the training is desirable.

ED's Even Start Program Provides Annual Training on Data Collection

For example, the ED's Even Start program provides annual training on data collection issues for new grantees and new data management staff who supervise local data collection and entry. As an example of the training activities covered, Even Start's training agenda for spring 1998 included

- orientation to the roles of various staff members and contractors involved with the data collection used for performance measures,
- discussion of the use and findings from similar data in prior evaluation reports to illustrate the importance of accurate data collection,
- directions and answers to frequently asked questions about six types of data collection forms,
- tips about local data entry methods and schedules, and
- demonstration sessions and opportunities for hands-on practice in using electronic data entry forms.

In addition to this annual training, evaluation and data collection concerns are discussed in meetings of grantees.

VHA Has Produced a Professional Training Video

To engage staff across the country in data quality issues, the Department of Veterans Affairs' (VA) Veterans Health Administration (VHA) produced a training video called "I am Joe's Data" for initial use at its Data Quality Summit. The video shows the "travel" of data from a typical patient in a VA hospital through various processing steps and illustrates the diverse ways in which the data are used. This video is being distributed to staff in VA hospitals to help them understand the importance of their roles in the accuracy of data that is ultimately presented to Congress.

Provide Feedback to Data Collectors on Types of Errors Found by Data Checks

When data quality checks are performed on information systems with ongoing data collection, agencies may provide feedback about the types and frequencies of errors found to those collecting and entering data. The feedback might list errors found in the data submitted by the specific data collection unit and provide comparison with error rates from other units or from all units. Sometimes the specific data entries with errors can be corrected; in other cases, obtaining accurate data in the future may be the objective. Feedback about data problems is sometimes combined with feedback showing the actual aggregated data results from that unit, so operating organizations see their concrete results along with any data problems.

ED's Even Start Contractor Provides Immediate Telephone Feedback on Errors

For example, the ED's Even Start contractor provides immediate telephone feedback about any data submission errors. Each project also receives a "project profile" that summarizes the data for its own project, compared with other similar projects, state averages, and national averages. The ED evaluation officer reported that such feedback contributes to data quality by encouraging projects to get their data in on time, with less data cleaning needed, and to be more involved in properly implementing any changes needed in the data collection procedures.

OPM Uses Agency Data to Verify Contractor Claims Processing Data

Another example of the use of feedback comes from the Office of Personnel Management, which oversees the life insurance program for federal employees. According to OPM staff, initial death claims processing is done by a contracted life insurance company, but OPM does a computerized "paid claims match," using agency records to verify the contractor's claims processing data. The results of these reviews are sent back to the contractor for investigation of any discrepancies, and results are fed back to the relevant managers within OPM. These data are also used in training new staff on the types of cases that may lead to errors in the adjudication of claims.

Use Analytic Methods and Transformations Appropriate for the Data Type and Measure Being Reported

Before reporting data as performance measures, it is often necessary to aggregate data from multiple locations, to transform the raw data into a ratio or percentage, or otherwise process the data.

DOT Measures the Rate of Fatalities per Vehicle Miles

For example, when reporting its performance measure for highway fatalities, the Department of Transportation (DOT) uses the rate of highway-related fatalities per 100 million vehicle miles traveled rather than the “raw” number of fatalities. This ratio adjusts for a greater risk of fatalities each year due to an expected approximately 2.2-percent annual increase in miles driven.

VHA Plans to Use Indexes to Report on the Quality of VA Health Care Delivery

An example of aggregating multiple data elements comes from the Veterans Health Administration, which plans to use several indexes, including the Chronic Disease Care Index and the Prevention Index, to report on the quality of its health care delivery. An index that includes information on a number of health areas allows the agency to provide an overall assessment of performance. VA’s fiscal year 2000 performance plan indicates that both indexes measure how well VA follows nationally recognized guidelines and recommendations for delivering clinical care to veterans with chronic diseases and for primary prevention and early detection. For each index, data about multiple relevant conditions are extracted from a sample of individual patient charts. The data are aggregated to form the indexes and statistically evaluated for validity and reliability.

An Alternative Approach to Performance Assessment at NSF

Not all agencies are depending entirely on using or building new quantitative data systems. The National Science Foundation is developing an alternative format for performance reporting that relies on qualitative assessments by external reviewers, as permitted under OMB’s Circular A-11 guidance. NSF’s procedures for these assessments illustrate some issues in verifying and validating qualitative methods to build quality into this alternative practice.

NSF is a federal agency that supports basic scientific research and science education. It operates primarily by awarding grants and cooperative agreements to individuals and groups in research institutions. For its alternative assessment approach, which is used for four major outcome goals on the advancement of science, NSF developed descriptive standards to characterize “successful” and “minimally effective” performance. For

example, for Outcome Goal 1: “Discoveries at and across the frontier of science and engineering,” the standards state the program is

- “successful when NSF awards lead to important discoveries; new knowledge and techniques, both expected and unexpected, within and across traditional disciplinary boundaries; and high-potential links across these boundaries” and
- “minimally effective when there is a steady stream of outputs of good scientific quality.”

Committees of External Reviewers to Assess NSF Research Grant Results

A committee of external reviewers for each scientific program will assess the program’s research grant results by applying these standards, using as evidence summary reports and examples of results prepared by program staff. Each committee’s report will be reviewed by several higher level entities: the Directorate’s chartered Advisory Committee of external scientists, the Directorate’s senior management, and the Office of the Director. These procedures build on similar prior peer review that focused primarily on improving the processes of grantmaking, which has been very useful as a management tool, according to an NSF official.

Validation and Verification Built in for NSF’s Alternative Assessment

NSF has built into its alternative assessment procedures several methods to increase the credibility of reports of program performance. First, NSF issued explicit guidelines on how the review committees will be convened and managed to help make the process systematic. Second, the guidelines require that the reviewers be “credible, independent experts who are able to provide balanced and impartial assessments,” with diversity among scientific, institutional, geographic, and demographic characteristics. Finally, the sequential layers of review for scientific programs help to validate the judgments made in the initial steps.

The external review assessments ultimately depend, however, on the selection of final project reports and other materials provided by the program staff to reviewers. NSF guidance does not require that the review include a balanced sample of projects closed out during the years being reviewed; instead, “examples may be selected to reflect the most significant accomplishments in a program’s portfolio of support.” Agency officials report that the reviewers will have access to all information systems and will be encouraged to make their own choice of examples. NSF intends to review this process and make changes.

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Order by mail:

**U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013**

or visit:

**Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC**

Orders may also be placed by calling (202) 512-6000 or by using fax number (202) 512-6061, or TDD (202) 512-2537.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touch-tone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

<http://www.gao.gov>

**United States
General Accounting Office
Washington, D.C. 20548-0001**

**Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100**

**Official Business
Penalty for Private Use \$300**

Address Correction Requested

