

STRONGEST AND WEAKEST SCORES

THIS SECTION LISTS the agencies, the performance and accountability reports, and the pilot reports that scored the highest and lowest for each of our 12 evaluation criteria. We briefly explain what made these reports the strongest or the weakest. The discussion of each criterion begins by describing the key factors we look for under that criterion and why we consider them important. Each ends by highlighting the main improvement opportunities we would offer based on our review of the reports this year.

TRANSPARENCY

1. Is the report easily identified and easily accessible online?

ACCESS TO PERFORMANCE information is critical because public accountability can only be served if members of the public can actually determine the benefits an agency produces. The annual report should be easily available to Congress, the public, stakeholders, and the media. Ideally this means that the agency's home page displays a link clearly guiding the reader to the annual report for the most recent fiscal year. If one has to be an expert on performance management, agency structure, or the agency's Web site to locate it, the spirit of accountability to the public is not satisfied. If the report is large, it should be divided into sections for more convenient reading and downloading. Making the report available in multiple formats is also desirable since readers' needs vary and each format has its advantages and disadvantages (ease of printing, searching, etc.). Finally, the agency should include contact information so that people can mail, phone, or e-mail questions, comments, or requests for a hard copy of the report.

We apply the following key evaluation factors for criterion 1:

- Did the agency post the report to its Web site on a timely basis?
- Is the report easy to locate on the Web site?
- Is it easy to navigate online and download?
- Does the agency provide prominent and useful contact information?
- Are prior-year versions of the report readily available?

Strongest Scores: Thirteen agencies (5 points)

MERCATUS RESEARCHERS REGARDED a PAR as "timely" if it was posted by December 1, and we regarded a citizens' report as "timely" if it was posted by January 19. Most agencies satisfied the basic elements of criterion 1 by posting their reports on time, creating a direct home page link to the report, permitting downloads in both single and multiple files, and providing at least some contact information. With one exception, the reports were posted on time. Indeed, few agencies needed until December 1 or January 19. Ten of the 15 PAR agencies posted the reports on their submission date, November 17, and 2 more agencies posted their reports within a day or two after November 17. All were posted by our December 1 deadline. Likewise, 8 of the 9 pilot agencies posted their citizens' reports to their Web sites either on the submission date of January 15, 2009 or the following day.

Weakest Score: SBA (2 points)

THE ONE EXCEPTION to timely Web posting was the SBA, which failed to post its citizens' report until well after the Mercatus deadline of January 19. The SBA also failed to post its performance report in time for us to consider it in our evaluation. Finally, the SBA citizens' report failed to identify, reference, or link to the inspector general's presentation on major management challenges in its fiscal year 2008 financial report. This is particularly unfortunate since the SBA inspector general's presentation is one of the most insightful of all such presentations.

Opportunities for Improvement

Make home page links more prominent. Two of the 15 agencies filing PARs lacked a direct home page link to the report. As in past years, the prominence of the links varied considerably. Some agencies (e.g., Interior, Justice, and Labor) highlighted the report via a prominent home page link. Others (e.g., NASA and SSA) had a relatively obscure link in small print at the bottom of their home pages. All nine pilot agencies had a direct home page link to the citizens' report. A number of the pilot agencies had prominent links, including Defense, HHS, OPM, and USAID. In a new feature this year, two agencies (NRC and USAID) included an introductory video by the agency head on the Web page that contains the citizens' report.

Provide better contact information, particularly online. As in past years, most agencies provided contact information for readers interested in obtaining paper copies of the report or asking questions about it. However, the usefulness of the contact information varied. EPA had particularly good contact information in its report. The report's table of contents page and the inside back cover identify extensive report-specific contacts. In addition, an e-mail address for submitting comments or questions about the report appears at the bottom of each page. Some agencies had no contact information online, forcing the reader to peruse the body of the report for such information. In some cases, the online contact information was not specific to the report. A number of agencies (e.g., HUD and Treasury) had online general contact information, such as an "About Us" link, for personnel in the office that produced the report but not necessarily those officials specifically knowledgeable about the report. In our view, the best practice is to post contact information online that puts the reader directly in touch with agency officials who have specific knowledge of the report. Examples of agencies having online, report-specific contact information are Education, Energy, Labor, and Treasury.

Improve links to prior-year reports. Prior-year performance reports for most agencies could be found online, but the ease of accessing them varied. In some cases, it took a series of searches to access the prior reports. For a number of agencies, the Web pages containing their current reports readily produced or linked to the prior-year versions as well: Commerce, Energy, Homeland Security, HUD, Interior, NASA, NSF, OPM, SBA, SSA, and Veterans Affairs. This is a best practice that all agencies could easily emulate.

2. Is the report easy for a layperson to read and understand?

The report should focus on what the agency's mission is, how the agency organizes efforts toward that end, and how much progress was made. The report's contents should be clear, logical, easy to navigate, and presented in such a way that the structure aids understanding. Consistent format, clarity of text, absence of jargon, and effective use of visual techniques like headings, graphs, tables, and photos are helpful. Acronyms can be helpful if they substitute for lengthy proper names with which readers may be familiar, but using acronyms to refer to documents, processes, systems, nouns other than proper names, or names of things known only to insiders inhibits understanding even if the report provides a list of acronyms. Details can either inform or confuse, depending on how they are presented. Anecdotes can promote effective communication if they complement

and illustrate, rather than substitute for, outcome-based results measures.

We apply the following key evaluation factors for criterion 2:

- Is the report written in plain language that a lay reader can understand?
- Is it well organized and easy for the reader to navigate and identify key information?
- Does it use clear presentational formats that the reader can readily comprehend, including effective tables and graphics?
- Are the narratives clear and concise?
- Does the report feature a manageable number of performance goals and measures that capture outcomes whose significance is apparent to a lay reader?
- Does it limit the use of technical terms, jargon, and acronyms?

Strongest Scores: Justice, Labor (5 points)

BOTH JUSTICE AND Labor produced highly readable PARs, which were improved by the addition of excellent citizens' reports. The key portions of the Justice report are much shorter than most PARs. At 22 pages, the Management's Discussion and Analysis section is the shortest of all reports, and the detailed performance section is a relatively modest 54 pages. The report is well formatted and lays out the department's performance results clearly. Justice uses a manageable number of performance indicators: 25 key measures with a total of 33 targets. The report includes reader-friendly tables that present at a glance the department's goals, performance measures, and fiscal year 2008 results.

Justice's citizens' report obviously was prepared with the general public in mind. Its sections have captions such as "The Department of Justice at a glance" and "How do we serve you?" It has excellent, user-friendly links to other source information. It includes the department's "snapshot" as well as summaries of performance results by strategic goals that are concise and well suited to a lay audience. The report features representative performance goals, which, while not consistently outcome oriented, are clear and focused on subjects of obvious importance to the public. It highlights accomplishments in each strategic goal area and includes vignettes to illustrate some of them.

Labor's report is also reasonably short by PAR standards. It is visually appealing and makes effective use of tables and graphics. The secretary's transmittal letter highlights the department's key achievements over time as well as fiscal year 2008 results for each strategic goal area. Each strategic goal presentation begins with an overview of the department's role and the public benefits it provides. The report includes a number of vignettes that illustrate the impact of the department's work on individuals.

Labor's citizens' report includes the department's "snapshot" and uses a shorter version of the secretary's transmittal letter from the PAR that retains the content on performance results but omits the technical content. The report is concise and well organized. It incorporates a number of the strong presentational features from the full report and has user-friendly direct links to other information sources. The main body consists of reader-friendly, insightful summaries of the department's results by strategic goal, using representative and primarily results-oriented performance measures.

TMA (Too Many Acronyms)

In past years, we've been treated to acronyms like LUST (Leaking Underground Storage Tanks) and GWOT (Global War on Terror). These acronyms caught our eyes in the fiscal year 2008 reports.

Acronym	Meaning	Perpetrator
COOL	Country of Origin Labeling <i>Will terrorists think this is cool?</i>	Agriculture
NIPLECC	National Intellectual Property Law Enforcement Coordination Council <i>A very intuitive one. Not!</i>	Commerce
YRBSS	Youth Risk Behavior Surveillance System <i>Sounds like it might be useful in the GWOT.</i>	Education
STARS	Standard Accounting Reporting System	Energy
STRIPES	Strategic Integrated Procurement Enterprise System <i>What a cute pair!</i>	Energy
WHAT If	Watershed Health Assessment Tools Investigating Fisheries <i>Does it include any "What if?" questions?</i>	EPA
IDAMWIG	Identify and Access Management Working Group <i>We wiggled out over this one!</i>	GSA
SPITS	Service's Permit Issuance and Tracking System <i>And what's the output from this system called?</i>	Interior
AFF/SADF	Assets Forfeiture Fund and Seized Asset Deposit Fund <i>Sounds sadistic.</i>	Justice
DOLLAR\$	Department of Labor Accounting and Related Systems <i>Might lead to worker's comp injuries when we stretch to reach the \$ sign on the keyboard.</i>	Labor
CAROLFIRE	Cable Response to Live Fire <i>Is this an instruction or a condition?</i>	NRC
PAAT	Performance Appraisal Assessment Tool <i>Useful for giving employees a PAAT on the back.</i>	OPM
SO	Strategic Objective <i>Almost takes longer to pronounce the acronym than the words.</i>	SSA
DASHR/CHCO	Office of Deputy Assistant Secretary of Human Resources/Chief Human Capital Officer <i>Is Treasury's Bureau of Acronyms under this officer?</i>	Treasury
SMART	Standardized Monitoring and Assessment Relief and Transition <i>Acronyms can be smart as long as you avoid the dumb ones.</i>	USAID
HIS	Indian Health Service <i>Obviously a victim of Microsoft Word's overactive spell-checker.</i>	Veterans Affairs

Honorable mentions:

HHS, for using nine acronyms in the one-page transmittal letter for its citizens' report.
OPM, for using "CR" as an acronym for its citizens' report.

Weakest Scores: HUD, SBA (2 points)

THE HUD REPORT is over 400 pages long and consists primarily of text, with only a few tables interspersed among its lengthy narratives. The report covers 6 strategic goals, 27 strategic objectives, and more than 100 individual performance measures. The detailed content for each strategic goal starts with a “performance scorecard” listing the performance measures and results for each measure applicable to that goal. However, the reader must delve into the following text to discover what some of the measures mean. The secretary’s transmittal letter is brief and says little substantive about the department’s performance. To its credit, HUD voluntarily produced a citizens’ report to complement its PAR. However, this document has little content on the department’s specific performance results.

SBA, a pilot agency, has a citizens’ report that has virtually no useful content within the document and lacks useful links to other information sources. While brief, it contains little information about the agency’s program performance. There is no agency head transmittal letter. Few specific performance measures are described. There are no accompanying narratives to explain the agency’s performance on the few measures that are included. Less than half of one page of the report is devoted to describing fiscal year 2008 accomplishments even in general terms. It does not mention the major management challenges identified by the inspector general and does not reference the inspector general presentation in the financial report. In sum, it falls far short of providing the overview of agency performance that a citizens’ report is supposed to convey. While SBA’s fiscal year 2008 performance report may supply much of the missing content, it was not accessible online at the time of our review.

Opportunities for Improvement

Shorten the PARs. A basic challenge for PARs is to convey information to the general public as concisely as possible through a document that has much required (and often highly technical) content and which must serve audiences with different needs. Size certainly can be a barrier to readability. Given its scope, the PAR will always be a fairly large document. However, the fact that PARs vary greatly in size indicates that agencies have considerable flexibility and that the reports need not be oppressively long. PAR agencies can (and do) use many techniques to kill fewer trees and package information more concisely. One common practice to reduce report length and enhance readability is to use fewer performance measures. This is a good approach for all agencies, PAR and pilot alike, and is discussed further below. Another technique, discussed below, is to use tables and graphics to summarize performance results. This saves space and enhances readability. Also, the reports might omit some technical content (e.g., data verification and validation details) in favor of providing a link to it online for those readers who are interested. Above all, report drafters should be less verbose. For example, it is hard to believe that the Management’s Discussion and Analysis sections of the reports, which are supposed to be overviews, need to consume an average of more than 50 pages as they now do. It also is hard to believe that descriptions of major management challenges need to consume over 80 pages (e.g., Transportation).

Make the citizens’ reports of pilot agencies more substantive and user friendly. While the PAR reports contain too much, we found that the citizens’ reports of pilot agencies often contained too little. For a pilot agency, the citizens’ report is the key document to inform the general public. It needs to serve as a self-contained, substantive overview of the agency’s performance and provide user-friendly links to other source documents for more detail. Three of the pilot citizens’ reports (Defense, OPM, and SBA) fell well short of this standard. The others were better, but could be improved. In particular, all need more user-friendly links directly to the relevant portions of their agency performance and financial reports. The Justice and Labor citizens’ reports serve as models for the pilot agencies both for their content and their direct links to other sources.

The following suggestions to improve readability apply to both PARs and pilots.

Keep the general public in mind. Agencies face many challenges in crafting their performance reports, but writing a report that ordinary citizens can read and understand should not be one of them. Since 1998, federal agencies have been obligated to write documents in “plain language” with logical organization, everyday words, active voice, short sentences, and easy-to-read tables and graphics, among other features.²² Report writers should consider how well their product will communicate to lay readers. It is easy to tell which reports were or were not written from this perspective. Reports may need technical content to satisfy the legitimate interests of specialized audiences. However, there are ample opportunities to reach the general public as well. Good venues for this are the citizens’ report, the transmittal letter, and the Management’s Discussion and Analysis section of the PAR. As discussed previously, Justice’s citizens’ report provides an excellent example of a document written for the general public. By contrast, the performance section of the GSA report offers an example of how not to write for the public. It is organized primarily by the agency’s components rather than by its strategic goals. This organization may make it easy for agency insiders to use the report, but it makes the report disjointed and difficult for the general public to navigate.

Focus on a clear and concise set of performance metrics. Some agencies have been quite successful in presenting their performance through a limited set of measures that capture important public outcomes. For example, Justice uses a total of 25 performance measures; Agriculture gets by with 33 measures. Agencies that find it necessary to have many more performance indicators can ease the reader’s burden by focusing their reports on a smaller set of key representative measures. Agencies that accomplished this include Commerce, GSA, Homeland Security, Interior, and Veterans Affairs.

Minimize lengthy, text-heavy narratives and maximize tables and graphics. Many reports make effective use of tables, vignettes with photographs, graphs, and similar features to highlight aspects of their performances or drive home key points. One particularly helpful feature is a “report card” or “scorecard” table that summarizes the agency’s overall performance at a glance. Examples are Veterans Affairs and SSA.

3. Are the performance data reliable, credible, and verifiable?

THE REPORTS CONSOLIDATION Act of 2000 (Public Law 106-531) provides that the agency head transmittal letter must include “an assessment by the agency head of the completeness and reliability of the performance and financial data used in the report” and that the assessment must “describe any material inadequacies in the completeness and reliability of the data” as well as what the agency can do and is doing to resolve them.²³

Like a similar requirement imposed on the chief executive officers of publicly held companies, the purpose of this requirement is to ensure that the quality of the underlying data is a management priority and that deficiencies are acknowledged and corrected as quickly as possible. More generally, the report should indicate the agency’s confidence in the quality of the data used to document its results. Data sources and descriptions should be provided for all performance measures.

22. See President Clinton's *Memorandum on Plain Language in Government Writing* (June 1, 1998), <http://www.plainlanguage.gov/whatisPL/govmandates/memo.cfm>, and "How to Comply with the President's Memo on Plain Language" (July 1998), <http://www.plainlanguage.gov/howto/guidelines/PresMemoGuidelines.cfm>.

23. These requirements are codified at *U.S. Code* 31 § 3516(e).

We apply the following key evaluation factors for criterion 3:

- Does the report fully and candidly assess data quality?
- Does it indicate that the agency has major data challenges?
- If so, does it provide confidence that the challenges are being addressed?
- Does the report include at least preliminary results for all measures?
- Does it provide data definitions and sources for its performance measures?

Strongest Scores: Labor, Veterans Affairs (5 points)

THE DATA QUALITY Act obligates federal agencies to ensure and maximize “the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by the agency. . . .”²⁴ The Labor report systematically assesses and rates the quality of its data for each performance goal, a best practice for other agencies to emulate. While the report identifies some concerns, it states that performance data are accurate and relevant for all goals. The performance section elaborates on data quality for each goal and provides data sources for each measure.

The Veterans Affairs report includes a detailed discussion of data quality, which has been expanded this year to include more specifics on verification techniques. Tables in the performance section provide detailed background on the data supporting the key measures, including data definitions, data validation and verification, and any data limitations. Additional data definitions and background appear later in the report.

Weakest Scores: Defense (1 point); Eight agencies (2 points)

THE DEFENSE REPORT has no content on performance data. The transmittal letter does not include the agency head data assessment required by the Reports Consolidation Act, and the body of the report does not address data quality in any way. The department received a disclaimer of opinion on its fiscal year 2008 financial statements. A disclaimer of opinion means the financial statements were not complete or accurate enough to allow the auditors to judge whether the statements fully and fairly present the agency’s financial position.

Many reports lack sufficient content to give the reader confidence in their supporting data. Some reports’ contents (or lack thereof) seem to contradict assurances that data are complete and reliable. The HHS secretary’s transmittal letter states that the report’s financial and performance data are reliable, complete, and the latest available except where otherwise noted. However, data completeness obviously is a major issue for HHS, because its report indicates fiscal year 2008 results are available for only 40 percent of the department’s performance targets overall and only about one-third of the measures featured in the report. The transmittal letter for the OPM report does not include the agency head data assessment required by the Reports Consolidation Act. Instead, the agency’s “performance improvement officer” gives a data assessment in the transmittal letter, stating that the performance information contained in that report “is reasonably complete and reliable, as defined by the Government Performance and Results Act.” However, the report later cites questions about the validity of two of its own measures as a reason for failing to meet them.

24. Cited in Office of Management and Budget, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, January 3, 2002, <http://georgewbush-whitehouse.archives.gov/omb/fedreg/reproducible.html>.

Opportunities for Improvement

MOST FEDERAL AGENCIES face serious challenges in developing complete and reliable performance data. This is particularly true of agencies that seek to achieve broad national outcomes that are difficult to measure and attribute, especially in annual increments (such as the results of multi-year basic research and development programs). Indeed, the more outcome-oriented an agency's performance metrics, the more difficult the measurement challenges may be. Another challenge confronting many agencies is the need to rely on third-party data sources. Recognizing that these challenges exist, our evaluations under criterion 3 place considerable weight on the agency's candor in disclosing data weaknesses and whether the report provides confidence that the agency is doing its best to address these weaknesses. From this perspective, two recurring problem areas call for improvement.

More candid agency-head data assessments. Unfortunately, many reports respond to the Report Consolidation Act's data assessment requirement with a perfunctory statement that the report's data are complete and reliable in accordance with OMB guidance or except as otherwise noted. Agency-head transmittal letters occasionally make such assertions even though the reports lack even preliminary current results for some measures or present what appear to be significant data reliability issues. This suggests that agency heads need to be more forthright, that the OMB guidance needs to be tightened up, or both. The EPA administrator's transmittal letter is one of the few that provides a candid and informative data assessment. He states:

Data used to report progress are reliable and as complete as possible. Inherent to the nature of our work is a time lag between when we take action to protect human health or the environment and when we can measure a result from that action. Therefore, for the reporting year, we cannot provide results data for several of our performance measures; however, we portray trend data, when possible, to show progress toward results over time, and we present final results for prior years when data have become available in fiscal year 2008.²⁵

Fuller data disclosure. Many reports simply do not provide enough information to give the reader a sense of how reliable the data are or even what data the agency uses. The credibility and usefulness of the report can be enhanced by providing data definitions and sources for individual measures. It is also helpful to describe what the agency does to verify and validate performance data. As noted above, Labor and Veterans Affairs are examples. Other examples of reports with good content on data verification and validation are Commerce, NRC, and NSF.

4. Did the agency provide baseline and trend data to put its performance measures in context?

THE "BOTTOM LINE" for citizens is whether an agency's actions are making a given situation better or worse—by how much and at what cost. To provide this information, agencies must design measurement systems that facilitate analysis of trends over time. Data should be displayed in a way that allows readers to detect and understand their significance easily. Both quantity of data (years of data included) and presentation matter. Good performance measures that have limited data (due to newness or revision) may convey more information than inferior measures with more data points that are not clearly linked to an agency's results.

Multiple years of data help identify trends, but they do not by themselves show how close the agency is to

25. Stephen L. Johnson, "Administrator's Letter", *U.S. Environmental Protection Agency Performance and Accountability Report Fiscal Year 2008* (Washington, D.C.: EPA, 2008), vii.

achieving its goals, or explain why the agency will produce a significant level of public benefits if it hits its targets. A report should explain rationales behind the selection and alteration of quantitative targets so readers can understand the magnitude of the agency's goals in relation to the size of the problem.

We apply the following key evaluation factors for criterion 4:

- Does the report provide prior-year results for comparable measures so that the reader can identify performance trends and track progress from year to year?
- Does it describe long-term goals so that the reader can assess the extent of progress?
- Does it explain the rationale for annual targets, particularly where there are significant differences between current targets and prior-year results?
- Do the trend data indicate that the agency is setting challenging targets in relation to past performance?
- Do the report narratives assist the reader by analyzing performance results and trends?

Strongest Scores: Labor, Veterans Affairs (5 points)

THE PERFORMANCE SECTION of the Labor report includes baseline and trend data for each measure generally going back four fiscal years or program years. These data include the prior-year targets, actual results, and costs. The report uses graphics to show further trend data for some goals and measures. In a best practice, the performance section includes for each goal a "Program Perspective and Logic" discussion, which provides valuable context to assist the reader in understanding the department's performance.

The Veterans Affairs report is rich in baseline and trend data and does a good job of integrating these data into its presentations. Performance summaries in the report show prior-year performance targets and results for key measures and representative supporting measures, generally going back to fiscal year 2004. The tables in the performance section show the strategic (long-range) target for each measure. This provides the reader with valuable context to assess the extent of the department's progress and how far it needs to go.

The Transportation and Treasury reports also feature good presentations of baseline and trend data. The Transportation report includes a performance "forecast" for each measure for fiscal year 2009. The narratives in the performance section elaborate on the department's results, adding valuable context for assessing its performance. In an innovative feature and potential best practice, the Treasury report uses graphics to show the trends for target achievement and actual performance over time.

Weakest Scores: Nine agencies (2 points)

THE ENERGY REPORT'S presentations are an incomplete and confusing mix of prior-year targets and results. Many of the measures are technical, and the narratives do little to help a lay reader understand the measures or put the department's performance in context. NASA and NSF largely use non-quantifiable performance measures that afford little basis for assessing progress from year to year. The Commerce and SSA reports have baseline and trend data suggesting that the agencies tend to set "lowball" performance targets; current targets are significantly below prior-year performance levels. There may be good reasons for this, but the reports do not explain them. SSA offers a particularly striking example. The administrator's transmittal letter describes

eliminating the hearing backlog for disability claims as a “moral imperative.” However, the backlog will not be eliminated until fiscal year 2013 under the agency’s new strategic plan.

Opportunities for Improvement

DATA SHOWING PRIOR-YEAR actual performance results for comparable measures provide the minimal information needed to identify performance trends and afford the reader some sense of how challenging an agency’s current targets are. The better reports have two additional features that make the baseline and trend data more insightful and put the agencies’ performances in fuller context.

Explanations for targets. One additional feature is an explanation of how performance targets are set. Explanations are particularly important where the trend data indicate that the agency’s current targets are not challenging in relation to past performance or where there are large fluctuations from year to year between targets and actual results. As noted above, the Labor report is strong in this respect. The Education report also includes a good “target context” feature.

Long-term perspective for performance trends. The second feature is providing a sense of each of the agency’s long-term goals in relation to its respective performance measure. Simple trend data can indicate whether the agency is making progress from one year to the next. However, it does not tell the reader where the agency hopes to go over the long term and thus enable the reader to assess the agency’s progress toward its destination. The Veterans Affairs and EPA reports do provide long-term strategic targets for this purpose. Several other agencies (e.g., Homeland Security and USAID) provide performance targets at least one year into the future.

PUBLIC BENEFITS

5. Are the goals and objectives stated as outcomes?

AN “OUTCOME GOAL” is defined as the intended benefit (or harm avoided) that results from an agency’s programs or activities. It should be articulated in clear terms that describe benefits to the community rather than activities that are presumed to be of value. For example, the number and percentage of children learning to read at grade level is an outcome; the number of children enrolled in a remedial reading program, or the number of schools receiving grants to fund these programs, are measures of activity, not outcomes. Vague verbiage that emphasizes things an agency does instead of why it is doing them should be avoided. This admonition applies at all goal levels—strategic goals, strategic objectives, and annual performance goals.

Strategic goals should be few in number. Management goals (financial, human resource, information technology, etc.) exist to support the achievement of genuine outcome goals. A majority of agencies now include a strategic goal focused on management improvement, but these should not be confused with actual programmatic outcome goals.

Criteria 5, 6, and 7 are closely related and all are crucial for effective performance reporting. It is difficult to achieve a high score on criterion 7, which assesses whether the report has demonstrated that the agency has an effect on outcomes, if the report has not articulated clear outcomes (criterion 5) and outcome-oriented performance measures (criterion 6). Similarly, it is difficult, though not impossible, to earn a high score on criterion 6 without a high score on criterion 5.

We apply the following key evaluation factors for criterion 5:

- Are strategic (long-term) and annual performance goals expressed as measurable outcomes that a lay reader can understand and relate to public benefits?
- If strategic goals are expressed at a high level of generality, are they accompanied by strategic objectives that provide more specific and measurable outcomes?
- Do the goals cover all of the agency's core missions?

Strongest Scores: Labor, Transportation (5 points)

WHAT MAKES THESE two reports stand out is that the great majority of the goals and objectives are stated as clear, measurable outcomes of obvious importance that the general public can readily understand and appreciate. For example, Labor's goals focus on such outcomes as safer workplaces and employment training programs that result in trainees obtaining and retaining jobs. Transportation's first strategic goal emphasizes increasing the safety of highway, aviation, and rail transportation by reducing accidents, deaths, and injuries. Its second goal aims at increasing transportation efficiency by reducing highway congestion and aviation flight delays. Other reports featuring strong outcome-oriented goals are Education, Health & Human Services, Homeland Security, NRC, and Veterans Affairs.

Weakest Scores: Six agencies (2 points)

THE WEAKER REPORTS lack goals with clear, measurable outcomes. Some have strategic goals, but they refer to outcomes at too high a level for the reader to understand how and what the agency contributes. Commerce's first strategic goal, for example, is to "Maximize U.S. competitiveness and enable economic growth." Reports frequently obscure the agency's role by using vague terms, such as "facilitate" or "enable." These reports also are short on results-oriented strategic objectives or annual performance goals that might flesh out the measurable outcomes.

Opportunities for Improvement

SCORES FOR THIS criterion have changed little from last year. Since few agencies implemented new strategic plans for fiscal year 2008, their strategic goals and objectives, as well as the annual performance goals, remained largely the same. This is not surprising given that this was the last year before a transition of administrations. One agency that did significantly change its strategic goals and objectives this year was Education, which made major improvements to the clarity and outcome orientation of its goals and objectives. Also, while retaining the same strategic goals as last year, State added descriptions to them for fiscal year 2008 that specify more clearly the measurable outcomes envisioned for each goal. State added text to its strategic objectives as well to clarify its intended outcomes.

The coming of a new administration offers an opportunity for agencies to develop new strategic plans and, in the process, improve the clarity and outcome orientation of their goals. Any agency should be able to develop a set of long-term strategic goals that capture the important public outcomes the agency exists to achieve in a clear and specific way that an ordinary citizen can grasp. Some of the agencies that score poorly here could improve by breaking down their current high-level strategic goals into more specific and measurable outcomes that could be tied with some credibility to the agency's contributions.

It is challenging for agencies to develop performance goals that convert outcomes into annual measurable increments. One approach that some agencies use is to include intermediate outcome goals. For example, increased use of seat belts is an intermediate outcome that promotes the end outcome of fewer highway deaths and injuries. It is a good intermediate indicator as long as the agency has factual evidence that increased use of seat belts does indeed lead to the desired outcome. Not all annual performance goals need to be outcomes or even intermediate outcomes. Output, activity, and efficiency goals have their place and may be important to some stakeholders. However, agencies should strive to have at least a readily identifiable set of outcome-oriented performance goals in order to demonstrate the agency's value to the general public in a credible way.

6. Are the performance measures valid indicators of the agency's impact on its outcome goals?

PERFORMANCE MEASURES SELECTED by an agency should relate directly to its outcome goals. Activity measures, such as number of participants, studies completed, facilities built, or projects funded, may contribute to achievement of a result but do not constitute results of interest to the public at large. Including these measures in the report may actually detract from the report's effectiveness in demonstrating the agency's impact. Data measuring levels of output can support a claim for success, but only if the agency makes a compelling case for a causal link between the output and results achievement.

We apply the following key evaluation factors for criterion 6:

- Do the measures and targets relate directly to outcome goals, and are they expressed as outcomes?
- Where measures are not stated as outcomes, do the accompanying narratives explain their relevance to mission outcomes?
- Where measures and targets are technical or otherwise unclear to a lay reader, do the accompanying narratives explain their significance?

Strongest Scores: Labor, Transportation (5 points)

NOT SURPRISINGLY, THE Labor and Transportation performance measures are highly outcome oriented since both reports have strong outcome goals. Labor's performance measures track well with their respective goals and are highly outcome oriented for the agency's strategic goals relating to a prepared workforce, safe and secure workplaces, and strengthened economic protections. The measures applicable to the competitive workforce strategic goal also include strong outcomes along with activity, output, and efficiency-type measures. All of the Transportation measures relating to the agency's programmatic strategic goals for safety, congestion, and security capture end or intermediate outcomes. At least half the measures for the global connectivity and environmental stewardship strategic goals focus on end outcomes or intermediate outcomes.

Weakest Scores: Eleven agencies (2 points)

THE COMMON THREAD running through the lower-scoring reports is that most of their performance measures address activities, outputs, processes, and efficiencies rather than outcomes. For example, although State made its goals more outcome oriented this year, its measures have become less outcome oriented. By our count, only 2 of the 16 programmatic performance measures featured in the State report capture outcomes. Most are raw

number measures. Many of the numbers deal with activities or outputs, such as number of election observers trained with U.S. assistance, number of host-country officials trained in counterterrorism, or number of foreign military personnel trained. Energy and NASA compound this problem by using arcane activity measures that are hard for a lay reader to understand, such as Energy's measure of "Efficiency of 'white light' solid-state lighting in a lab device, in lumens per watt (LPW)" and NASA's "Design, build, and deliver for flight two Foton M3 experiments." Many agency measures focus heavily on outputs, such as levels of assistance provided, but fail to capture what public benefits the assistance achieves. HUD and SBA are examples. Efficiency, workload, productivity, and customer satisfaction measures may be valuable for certain purposes, such as internal agency management. However, they tell the public little about outcomes.

Opportunities for Improvement

Convert raw numbers measures into outcome form. Many agency measures have the potential to capture outcomes but are not stated in an outcome-oriented manner. Raw-number measures are a recurring case in point. They may relate to outcomes but are expressed in a way that only measures activity levels. For example, one highlighted measure used by Homeland Security is the number of border miles under effective control. Border security certainly is an outcome. However, simply reporting on the number of border miles under control gives the reader little sense of progress toward that outcome. The measure would be more meaningful and outcome oriented if expressed as a percentage of border miles under control. There are many opportunities, particularly in law enforcement areas, to convert raw-number measures into percentages or other forms that capture progress toward the desired outcome. For example, such measures as the number of organized criminal enterprises dismantled or quantities of illegal substances seized do not tell the reader whether the outcome-related conditions they address (prevalence of criminal gangs or illegal drugs) are getting better or worse.

Eliminate measures lacking external significance. In addition to developing more outcome-oriented measures, agencies might consider eliminating from their reports measures that have little if any relevance to the public or other external stakeholders. For example, workload, productivity, and other process-related measures seem relevant primarily, if not exclusively, to agency officials for internal management purposes. Such measures contribute only to the size of the performance report. Customer satisfaction measures also fall into this category. Indeed, they are sometimes treated as a substitute for true outcome-oriented measures that express more directly the degree to which an agency is meeting the public's needs.

7. Does the agency demonstrate that its actions have actually made a significant contribution toward its stated goals?

THE REPORT SHOULD show whether things improved because of what an agency did, and if so, how much of the improvement can be attributed to its actions. Claims of impact should be supported by program evaluations or other empirical evidence of a cause-and-effect relationship. A less desirable alternative would be to logically connect outcome measures, output measures, and anecdotal evidence. A case that rests on merely assumed cause-and-effect relationships is unsatisfactory. The report should explain how agency outputs create or enhance outcomes for the public and describe the nature and extent of influence so that outcomes can be attributed (at least in part) to specific agency actions.

Discussion of the operating environment and the extent of the agency's influence are helpful in keeping expectations ambitious, yet realistic. External factors, however, should be treated as influences that must be controlled for in order to identify the agency's contribution—not as excuses for a failure to demonstrate performance.

We apply the following key evaluation factors for criterion 7:

- Does the report have overall performance metrics, strong goals (criterion 5), and strong measures (criterion 6), which combine to provide a clear basis for assessing agency performance toward end outcomes?
- Do the narratives in the performance section help the reader understand the impact of the performance results, particularly where their significance is not self-evident?
- Does the report provide sufficient current results, prior-year data, and other information to give the reader a complete picture of the agency's performance?

Strongest Score: Transportation (5 points)

THOSE AGENCIES WITH strong, outcome-oriented performance goals and measures will do well under this criterion, particularly if the narratives describing specific performance results are also strong. As noted previously, Transportation's performance metrics are predominantly results-oriented and focus on outcomes of obvious public importance. The narratives in the performance section complement the strong metrics and are highly effective in elaborating upon the department's performance results. Labor, NRC, and Veterans Affairs likewise scored well here based on their strong performance metrics and insightful descriptions of performance results.

Weakest Scores: Fourteen agencies (2 points)

JUST AS STRONG performance metrics largely determine success on this criterion, the predominant feature of the low-scoring reports is weak performance metrics—specifically, the absence of significant outcome orientations in the agencies' goals or measures. Of the 14 reports that scored a 2 here, all but one also scored a 2 for their goals (criterion 5), their measures (criterion 6), or both. The exception was GSA, which received a score of 2 on this criterion for incomplete reporting. In addition to weak performance metrics, some of the low-scoring agencies had weak descriptions of performance results (e.g., SBA); appeared to set non-challenging performance targets (e.g., Commerce and SSA); or had technical measures that were hard to understand (e.g., Energy and NASA). Three agencies (Energy, NASA, and NSF) constructed their measures in such a way that they say virtually nothing about value to the general public.

The absence of current (fiscal year 2008) results for a high proportion of performance measures is a significant limitation for four agencies: Education, EPA, GSA, and Health & Human Services. There is no apparent reason why GSA fails to report current results for many of its measures. For the other three agencies, the problem stems from data lags that may be largely unavoidable. Education and HHS deal with outcomes that take years to materialize and do not readily lend themselves to measurement in annual increments. To its credit, HHS mitigates this problem somewhat by reporting the most recent performance results available for those measures lacking fiscal year 2008 results.

Opportunities for Improvement

The improvements suggested in a number of other evaluation criteria will pay dividends here as well. The most directly relevant are criteria 5 and 6, of course, but criteria 2 through 4 also come into play.

8. Did the agency link its goals and results to costs?

WHEN RESULTS ARE linked to costs, decision makers know how much they have to pay per successful result. Armed with this information, they can assess the opportunities forgone when resources are allocated to less effective programs, and they can estimate how much more could be accomplished if resources were reallocated to the most effective programs.

This year's scoring roughly corresponded to the following standards: 1 point if the report had no relevant content or linked budget resources only to strategic goals, 2 points if the linkage was carried down only to the strategic objectives, 3 points in cases where there was some useful content beyond the strategic objective level but short of the performance goal level, and 4 points if the linkage was carried down to the individual performance goals. To earn 5 points, the report needed to link budget costs to individual performance measures.

We apply the following key evaluation factors for criterion 8:

- Does the report link its budget resources to individual performance measures, thereby permitting cost-effectiveness analysis?
- Do the linkages include all budget resources, including indirect costs?

Strongest Scores: Labor, Transportation (5 points)

THE LABOR REPORT links costs to the strategic goals and performance goals for the current year and the two previous fiscal or program years. In addition, the performance section links costs for the current year and several prior years to most of the department's individual performance measures. Likewise, Transportation has links to its individual performance measures. Other agencies are making progress here by linking costs to a portion of their performance measures or by developing innovative techniques to relate costs to performance.

Weakest Scores: Eleven agencies (1 point)

THE REPORTS THAT scored 1 for this criterion either failed to link budget resources to their performance metrics at any level or provided linkages only at the strategic goal level.

Opportunities for Improvement

WITH RESPECT TO this criterion, agencies can be divided into two camps: those that are making progress toward linking costs to performance metrics in a meaningful way and those that are not. The former have scores of 3 or higher. The latter have scores of 1 or 2. For the agencies that fall into the latter camp, the most obvious improvement strategy is to at least start trying. Developing linkages to the point of permitting meaningful cost-benefit analysis presents methodological challenges. It may be impractical or even impossible for some agencies to allocate costs to every individual performance measure. The goal should be to find the right unit of analysis for the linkage so that the agency can assess performance results in relation to funding investments. Ideally, the agencies that are leading the way now will resolve or ameliorate the challenges and develop best practices for the rest to follow, and the rest will try harder in the future.

Some agencies are pursuing innovative techniques here that may provide a model for others. The Treasury report presents information primarily on the basis of "performance cost" as opposed to budgetary resources. As the report explains, this approach provides a more accurate picture of the actual costs related to performance

since it includes “imputed costs, depreciation, losses, and other expenses not requiring budgetary resources.” It also covers costs for programs funded by sources other than general-fund appropriations, such as user fee-financed or non-appropriated fund activities. Interior allocates costs to the 25 representative performance measures featured in its report and, through activity-based costing graphs, shows performance results in relation to funding levels. The narratives also provide some analysis of the impact of funding levels on performance. This innovative feature is a potential best practice for performance reports.

FORWARD-LOOKING LEADERSHIP

9. Does the report show how the agency's results will make this country a better place to live?

AGENCIES SHOULD REALIZE and articulate the value they provide to the country. Reports should speak directly to the public about how agencies produce benefits that are important to citizens. Just as the best corporate reports feature communication directly from the chief executives, agency reports should demonstrate accountability of agency heads for their organizations' performances. Lofty ideals must be supported by outcome orientation, sound strategies, and successful achievement discussions. The reports should create confidence in agencies' abilities to improve America's future. Anecdotes and success stories can be important communication strategies in this regard, but their value is limited if they are not backed up by solid performance data.

We apply the following key evaluation factors for criterion 9:

- Do the report narratives describe public benefits that flow from the agency's work in a way that lay readers can understand and appreciate?
- Does the report illustrate the agency's public benefits with specific outcome-oriented examples of accomplishments?
- Does the report back up its narratives with strong performance metrics that affirmatively demonstrate the public benefits the agency achieves?

Strongest Scores: Transportation, Veterans Affairs (5 points)

COMPLEMENTING ITS STRONG performance metrics, the Transportation report's narratives do an excellent job demonstrating the important public benefits that flow from the department's work. For example, the secretary's transmittal letter and the Management's Discussion and Analysis section both highlight major performance accomplishments. The secretary's transmittal letter (p. 3) notes that increased seat belt use in 2008, largely due to high-visibility enforcement campaigns, resulted in an estimated 270 lives saved. The performance highlights section (p. 18) states that the fatality rate from highway accidents has dropped more than 13 percent from 1998 to 2007. The Veterans Affairs report narratives do an excellent job of describing the department's accomplishments and the public benefits that flow from them. The secretary's transmittal letter highlights a number of specific accomplishments. The Management Discussion and Analysis section details the most important accomplishments for each strategic goal area. The personal vignettes further contribute in this regard. These and other narrative portions of the report are backed up by solid performance metrics.

Labor also scores well on this criterion. A vignette in the Labor report (p. 102) describes how a low-income widower at age 80 was able to obtain work through training provided under a department-funded mature workers program. Another (p. 111) concerns a disabled Iraq War veteran who was unable to return to his prior

civilian job but acquired enhanced information-technology skills through a department-funded training program. Through this training, he qualified for a new job teleworking from home. In yet another example (p. 141), the department funded grants in 20 states to use debit cards to make faster and more secure unemployment insurance payments to claimants who lacked bank accounts.

Weakest Scores: Defense, HUD, OPM, SBA (2 points)

DEFENSE, OPM, AND SBA are pilot-report agencies whose inadequate, bare-bones citizens' reports convey little insight into their performance results or public benefits achieved. The lengthy HUD report has some useful narrative descriptions of performance highlights and accomplishments. However, they do not compensate for the department's weak performance metrics. Indeed, a reader might wonder about the credibility of the department's metrics when it reports achieving over 85 percent of its targets in a year that featured a nationwide housing crisis. The secretary's transmittal letter has little substantive content in terms of performance accomplishments or strategies to address future challenges.

Opportunities for Improvement

Maximize opportunities to tell the agency's story to the public. The narrative portions of the report, which have less constrained formats than some other portions, offer many opportunities to address the general public. The higher-scoring agencies have found effective ways to take advantage of these opportunities. They use the transmittal letters as more than perfunctory documents that recite the necessary legalisms. The best transmittal letters highlight important outcome-oriented accomplishments. Examples include Agriculture, Interior, Labor, and Transportation.

Highlight accomplishments to illustrate performance results. The narrative portions in the better reports describe performance highlights in a way that the general public can understand and appreciate. They frequently include vignettes illustrating how the agency's programs have benefited specific individuals. The narratives can be particularly useful for agencies that have difficulty translating their performance accomplishments into clear and specific annual outcomes. This includes research and development agencies, agencies with highly technical missions, and agencies whose broad-outcome accomplishments take years to mature. In particular, NSF should describe more examples of public benefits that result from the projects it funds since its performance metrics provide little insight about effects on the public.

Improve performance metrics. Even the most compelling narratives are not a substitute for solid performance goals and measures. While an agency's performance metrics are covered primarily by other evaluation criteria, a report cannot earn a strong score in criterion 9 with weak performance metrics.

10. Does the agency explain failures to achieve its goals?

IF AN AGENCY cannot identify reasons for failure, its ability to claim credit for success is suspect. Successes and failures that really matter occur at the strategic goal and objective level. The report should aggregate performance goal results and assess their impact on high-level goals. These summaries should take into consideration the fiscal year's priorities and the relative significance of different goals, measures, and actual results. Transparency and accountability are ill-served by merely listing detailed measures and data from which the reader is expected to draw conclusions.

It should be clear why specific targets were chosen. What are the upper and lower limits of acceptable and achievable performance and why? The effects of unexpected events or barriers—both internal and external—should be explained and the solutions revealed or suggested. Special care should be taken with resource explanations to indicate precisely how more or different resources would fix the problem and why reallocations were not made internally.

We apply the following key evaluation factors for criterion 10:

- Does the report clearly and candidly disclose performance shortfalls?
- Does it explain shortfalls in a way that gives the reader useful insight into their causes?
- Does it describe planned remedial actions in a way that gives the reader confidence that the agency is working to improve its performance?

Strongest Scores: Veterans Affairs (5 points); NASA, Transportation, and Treasury (4 points)

THE VETERANS AFFAIRS report clearly discloses performance results, including shortfalls. It contains a specific “Performance Shortfall Analysis” section, which describes causes and resolution strategies for significant performance shortfalls. This is a useful feature and a potential best practice. Under the heading “How VA Leadership Uses Results Data” in the detailed performance section, the report briefly describes performance improvement steps even where targets were met. Likewise, a table in the NASA report, captioned “FY 2008 Performance Improvement Plan,” lists each measure that was not fully achieved, explains why, and outlines plans for achieving the measure in the future. The Treasury report specifies gradations of performance results, both in the percentage of achievement for individual targets and the level of success. Where targets were missed, an appendix consistently provides an explanation under the heading “Future Plans/Explanation for Shortfall.” The explanations are brief but usually informative, and they are accompanied by a general description of remedial steps. The Treasury report clearly discloses performance shortfalls and offers concise and generally informative explanations where targets were missed or are projected to be missed. The secretary’s transmittal letter candidly highlights several major performance concerns.

Weakest Scores: SBA (1 point), 12 agencies (2 points)

THE SBA REPORT provides no explanations for the four shortfalls shown, all of which missed their targets by substantial margins ranging from 20 to 35 percent. SBA’s citizens’ report does discuss economic factors influencing loan volume, but it does not link these factors to the missed targets. Most of the reports scoring 2 for this criterion had incomplete or weak explanations of performance shortfalls. For example, the Commerce report offered no explanation for two of the four key measures that were missed. The explanation for another missed key programmatic measure was vague and unsatisfying, especially considering that the shortfall involved highly publicized problems with a proposed automated census system. The Education report had weak explanations for the three shortfalls reported. One suggested that the measure was arbitrary; another suggested that the target was too ambitious; the third really provided no explanation. The explanations of shortfalls in the OPM report sometimes seem to raise more questions than they answer. For example, the report attributes one shortfall to the fact that the performance data used did not align with the measure. The explanation of another shortfall calls into question the validity of the measure.

Opportunities for Improvement

Explanations, not excuses. Explanations for shortfalls are most valuable as a diagnostic tool to identify their causes and devise improvement strategies. Too often, however, the explanations amount to no more than excuses that provide no useful insight. Common examples are statements to the effect that the target was challenging or too ambitious, the agency came close and will try harder next year or the shortfall was attributable to factors beyond the agency's control. These may sometimes be true, but if so, the report needs to explain how the agency will correct its planning process so that those kinds of mistakes do not recur. The most useful shortfall explanations provide specifics that point the way to future improvement.

Reporting perfect or near perfect performance is cause for skepticism, not celebration. No reasonable person would expect a major federal agency (or any organization with complex and inherently difficult missions) to have no performance shortfalls. Rather than assuming that the organization was performing flawlessly, the reader would more likely suspect that the agency either was not reporting candidly or was not challenging itself in its performance measures. NSF serves as an example: The agency reported no performance shortfalls with respect to its three programmatic strategic goals this year, nor has it reported any such shortfalls at least as far back as fiscal year 2002. However, NSF uses a performance assessment system that, while perhaps technically sound, is highly unlikely to surface any instances of unsuccessful or even marginal projects. In NSF's assessment system, outside experts base their reviews primarily on best examples of projects provided to them by NSF program managers, rather than using a random sample or another more comprehensive or representative slice of the projects the agency funds. Thus, evaluators are not routinely exposed to anything but the agency's best work.

11. Does the report adequately address major management challenges?

THE REPORT SHOULD describe how risks to an agency's success are being minimized so as to maximize results for citizens. The impact of management issues is clearest in a context of specific goal achievement. It should be clear which challenges are "mission critical" and why. Major management challenge discussions should include full disclosure of the background, comments of the agency's inspector general and the GAO, agency responses indicating an appreciation of threats to its mission and goals, and an anticipation of future risks.

The Reports Consolidation Act requires performance reports to include a statement by the agency's inspector general that (1) summarizes the agency's most serious management and performance challenges and (2) "briefly assesses the agency's progress in addressing those challenges."²⁶ These statements play an important part in our evaluations for criterion 11. As we have done in prior years, the research team also considered the agency's "status" scores on the President's Management Agenda.

The inspector general statements play an important part in our evaluations for criterion 11. We also carefully review the agency management's response, if there is one, to the statement. It is important for the agency response to identify any differences it may have with the inspector's assessment. The inspectors, of course, do not have a monopoly on truth, and agencies need not agree with their assessments of the challenges or their view of priorities. The reader is best informed by a complete and candid exchange between the inspector and agency management. We found several instances of differences between inspectors and agency management. While HUD management generally agreed with the challenges identified by its inspector general, it rejected

26. U.S. Code 31 §3516(d).

some aspects of the inspector's problem statements and explained why. For example, management disagreed that the department lacked an adequate database to implement a program for the redevelopment of abandoned and foreclosed homes (see pp. 378 and 386). Also, Treasury countered its inspector's assessment of progress in improving performance and financial data at the Internal Revenue Service (see pp. 257–258 and 268). The most striking example is the GSA report (pp. 186–188), in which agency management takes issue with most of the inspector general's assessment.

We apply the following key evaluation factors for criterion 11:

- Does the report include the inspector general's summary and assessment of the agency's most serious management and performance challenges?
- Does it include a response by agency management to the inspector general's presentation?
- Does the report provide confidence that the agency is taking specific and effective actions to resolve major management and performance challenges?
- Does the agency have good "status" scores on the President's Management Agenda?

Strongest Scores: Labor (5 points); EPA, Transportation, and Veterans Affairs (4 points)

LABOR'S INSPECTOR GENERAL does a good job of summarizing the challenges and assessing the department's progress toward resolving them. The department itself uses a "traffic light" system to assess its own progress on each challenge. The assessments are presented in tables that discuss each challenge, the actions taken in fiscal year 2008 to address it, and the remaining actions needed along with their expected completion dates. Both the inspector's assessments and the department's self-assessments provide confidence that the department is working seriously to address each challenge. Another source of confidence is that the department has achieved "green" ratings on each of the core items of the President's Management Agenda for the past two fiscal years.

The EPA report describes material weaknesses and other significant deficiencies identified by the OMB, the GAO, and the agency's inspector. Like Labor, EPA achieved "green" scores on all the core items of the President's Management Agenda. The Management's Discussion and Analysis section of the Transportation report breaks down the challenges identified by the department's inspector general into their component parts and uses a "progress meter" graphic to self-assess its progress toward resolving each component. The department reports "moderate" or better progress for most items. Ironically, departmental management's discussion of the challenges is much more specific and insightful than the inspector's verbose and at times confusing presentation. The Veterans Affairs report has a table showing each inspector-identified major management challenge, its underlying components, and the strategic goal to which it most closely relates. The table also gives an estimated time frame for resolution of each challenge—although many are open ended ("2009 and beyond")—and it provides a page reference for a detailed discussion of the challenge. The report uses the same format to describe high-risk areas applicable to the department as identified by the GAO.

Weakest Scores: Seven agencies (2 points)

THE FUNDAMENTAL COMMON shortcoming in the low-scoring reports is that they provide the reader with little or no confidence that the agencies are serving as good stewards of taxpayer dollars by taking effective steps to address their major management challenges. The reports often are long on statements of commitment to resolve the challenges and general assertions of progress, but short on specific descriptions of significant remedial

actions. This is a particular concern since many of the challenges agencies face have persisted for years. Defense is the leading example of a report that provides few specifics about remedial actions in the face of pervasive and long-standing management challenges. Inadequate inspector general assessments often contribute to the weakness of the reports in this area.

Opportunities for Improvement

Better inspector general assessments of progress. The Reports Consolidation Act requires reports to include a statement by the agency inspector general that not only summarizes the most serious management and performance challenges, but also assesses the agency's progress in addressing them. The reports consistently include the inspector general's descriptions of major management challenges. What is often missing, however, is the required assessment of progress, which Congress considers an important "barometer."²⁷ Too many inspector general statements are lengthy tomes that describe the challenges in great detail but offer little if any insight into what key steps are needed to remedy them, whether the agencies are taking those steps, and when they are likely to produce a solution. Indeed, agency management's self-assessments of the challenges and progress toward resolving them can be more informative (and more concise) than the inspector general's statement. Transportation is an example.

For a number of years, the SBA inspector general presentation has provided the best example of an insightful assessment that should serve as a model for other inspectors general. The SBA inspector's assessment breaks down each challenge into specific recommended actions to resolve it. It includes tables that use a color-coded system to rate the agency's status with respect to each action and list remaining actions needed to resolve it. It includes a table summarizing the agency's overall progress for each action and challenge. Other agency inspectors general are now starting to follow suit. Like SBA, the Homeland Security inspector uses a "scorecard" to assess the department's overall progress on four of the nine challenges identified as "Limited," "Modest," "Moderate," or "Substantial." His scorecard also breaks down the four selected challenges into components and rates the department's progress on the components as well. The inspectors for Education and Labor also provide good progress assessments, although they are not as specific as those of the SBA and Homeland Security.

More specifics from agency management. Of course, agencies also need to improve their own presentations on major management challenges and give more specifics on the remedial actions they are taking as well as the timing and significance of those actions. More specific inspector assessments probably would encourage the agencies to be more specific in their responses. A few agencies do not respond at all to the inspector general presentations; they need to do so.

12. Does the report describe changes in policies or procedures to do better next year?

THE INTENT OF the Government Performance and Results Act is not just reporting for its own sake. The law's intent is to hold agencies accountable for results rather than for activities. The idea is to gather information on results and then to use that information in a strategic manner, as a guide to future decisions. The most important improvement will therefore occur at the highest level, rather than in individual program goals or with the adjustment of measures. Is it evident that the agency is using knowledge gained from the reporting process to

27. S. Rep. no. 106-337 (2000), 4.

revise its priorities and guide its activities? What is the potential for an agency to make a positive difference in the future? How will it realize that potential?

We apply the following key evaluation factors for criterion 12:

- Does the report do a good job of describing improvement plans to address specific performance shortfalls and major management challenges?
- Does it describe how to do better even where current performance met expectations?
- Does it evidence a strategic focus by describing more general future challenges that the agency faces in accomplishing its missions and how to address them?
- Does it evidence efforts to enhance performance accountability and reporting?

Strongest Scores: Labor, Veterans Affairs

BOTH THE LABOR and Veterans Affairs reports contain thorough, substantive, and thoughtful descriptions of improvement strategies for specific programmatic shortfalls and major management challenges. They also do an excellent job of describing future challenges that transcend individual measures and how the department plans to meet them. Finally, both departments continue to be leaders in developing innovations and improvements with respect to performance accountability and reporting. The Transportation report shares many of the same features.

Weakest Scores: SBA (1 point); Seven agencies (2 points)

OUR REVIEWERS COULD find no content on this subject in the SBA report. The other reports with weak scores tend to have little insightful content on improvement strategies for specific programmatic or management shortcomings and little to say about the broader mission-related challenges facing the agencies or how those challenges will be addressed. Some are among the agencies that seem reluctant to set challenging performance targets. The HUD report is an example. One striking feature is the report's lack of content on the current housing crisis. By contrast, the Treasury report highlights the department's actions and strategies to address the current financial crisis.

Opportunities for Improvement

IMPROVEMENTS IN A number of other evaluation criteria would pay dividends here. In particular, agencies could benefit by enhancing their descriptions of remedial actions for performance shortfalls and management challenges. Innovative efforts to enhance the quality of the reports in any category would help here as well.