INTRODUCTION

"Nothing could be more irrational than to give the people power and to withhold from them information, without which power is abused. A people who mean to be their own governors must arm themselves with the power which knowledge gives. A popular government without popular information or the means of acquiring it is but a prologue to a farce or a tragedy, or perhaps both."

—James Madison[1]

The Government Performance and Results Act (GPRA)[2] celebrates its 15th anniversary in 2008. Though he surely had no inkling of GPRA, James Madison clearly articulated GPRA's fundamental rationale when he inked the above statement in 1822.

Enacted in 1993, GPRA directs federal agencies to define the outcomes their agencies seek to produce, identify measures that show whether they are making progress on these outcomes, and disclose the results to Congress and the public.[3] Federal agencies are supposed to be accountable to elected policy makers. Elected policy makers are supposed to be accountable to citizens. The GPRA-mandated disclosure of goals and results is a crucial step toward both types of accountability.

Ever since agencies issued their first performance reports for FY 1999, the Mercatus Center has assembled a research team to assess the quality of their disclosure. As in past years, this Scorecard assesses the reports by the 24 agencies covered by the Chief Financial Officers Act (CFO Act), which accounted for virtually all federal outlays in FY 2007.[4] The 24 comprise all the Cabinet departments, plus the largest independent agencies.

Starting in FY 2002, agencies could combine their annual performance report with an annual financial report to produce a "performance and accountability report." That year marked the beginning of a slow but steady improvement in average scores on our Scorecard, from 28.8 in FY 2001 to 36.4 in FY 2006. The climbing score understates the improvement since we tighten the scoring criteria over time to reflect new best practices in performance reporting that agencies have developed.

FY 2007 reversed that trend. The average total score fell by 5 percent, from 36.4 to 34.6. After 9 years of reporting, 35 percent of non-interest federal spending by the 24 agencies was covered by reports with total scores of "satisfactory" or better in our evaluation.[5] Covering $916 billion of non-interest expenditures, these reports came from 7 of the 24 Chief Financial Officers Act agencies, as Figure 1 shows. This figure does represent a

1. James Madison, Letter to W. T. Barry, August 4, 1822. In P. R. Fendall (Ed.), *Letters and Other Writings of James Madison, Published by Order of Congress*, Vol. 3 (Philadelphia: Lippincott, 1865), p. 276.
2. For text of legislation, see http://thomas.loc.gov/cgi-bin/bdquery/z?d103:SN00020:%7CTOM:/bss/d103query.html%7C.
3. For a brief account of the evolution of federal performance reporting and the laws that influenced it, see Maurice McTigue, Henry Wray, and Jerry Ellig, *7th Annual Performance Report Scorecard: Which Federal Agencies Best Inform the Public?* (Mercatus Center 2006), pp. 21-22, available at http://www.mercatus.org/publications/pubid.2265/pub_detail.asp.
4. The principal parts of government not included in these 24 agencies are the judiciary, the legislative branch, the executive office of the president, and the independent agencies not among the 24 CFO Act agencies. *See Budget of the United States Government, Fiscal Year 2009, Historical Tables*, Table 4.1, available at http://www.gpoaccess.gov/usbudget/fy09/hist.html. Outlays for these agencies actually exceed the "total outlays" figure, but they total 96 percent of total outlays minus undistributed offsetting receipts.
5. Reports can earn a maximum of five points on each of 12 criteria. A score of 3 on a criterion corresponds to a "satisfactory" rating, and a report with an average score of 3 across all 12 criteria would earn a 36. Thus, a report must earn a score of 36 or higher to be classified as "satisfactory." For further explanation, see pp.11 below.

Figure 1: Fiscal 2007 Spending ($Billions) Covered by "Satisfactory" Disclosure

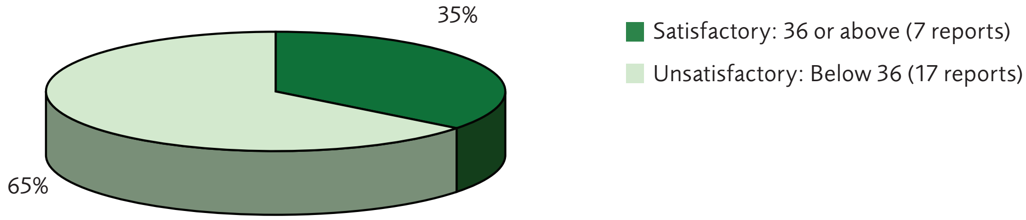
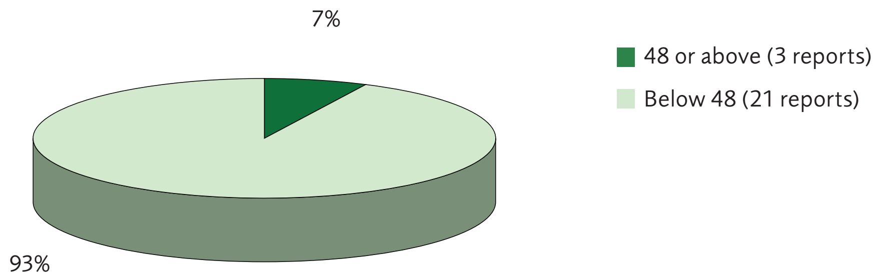


Figure 2: Fiscal 2007 Spending ($Billions) Covered by "Very Good" Disclosure (48 Points or Above Out of 60)



substantial improvement over previous years. In FY 2006, for example, only $337 billion (13 percent) of non-interest spending was covered by "satisfactory" reporting. Nevertheless, two-thirds of federal non-interest spending was covered by "unsatisfactory" disclosure.

The percentage of expenditures covered by "very good" disclosure was virtually unchanged from FY 2006.[6] Figure 2 shows that about 7 percent of non-interest spending, or $188 billion, was covered by reports scoring 48 or better. In FY 2006, 8 percent of non-interest spending was covered by "very good" reports.

Two main factors explain these changes. First, four of the agencies that saw large changes in their scores and rankings operated under new strategic plans for FY 2007. A major cause for the difference in their scores is improvement or deterioration in the outcome orientation of the goals and measures established in their new strategic plans. A strategic plan that articulates outcome-oriented strategic goals, objectives, and measures is a critical prerequisite if performance reports are to accurately disclose how the agency achieved outcomes of

6. A score of 4 on a criterion corresponds to a "very good" rating. A report with an average score of 4 across all 12 criteria would earn a 48. Thus, a report must earn a score of 48 or higher to be classified as "very good." For further explanation, see pp. 11 below.

value to the public. An outcome-oriented measure for a public health agency, for example, would be a measurable decrease in public health risk, such as fewer cases of a disease or its severity, as opposed to a measure of the number of firms inspected or a decrease in the level of contaminants.

Second, for FY 2007, the Office of Management and Budget (OMB) let agencies opt for a “pilot” format that allowed them to publish detailed performance information with their congressional budget justifications, required them to produce a separate financial report, and required them to produce a shorter “highlights” document intended to summarize both the performance and financial results. OMB explicitly stated that the purpose of the pilot was to allow agencies to explore alternative reporting formats that might improve the quality of information available to the public. Nine agencies chose the pilot format; 15 agencies chose to produce traditional performance and accountability reports. Of the latter, eight also chose to produce a highlights document even though OMB did not require them to do so.

The scores in this year’s Scorecard are one way to evaluate the pilot format. The pilot format has not improved the quality of performance information available to the public. For FY 2007, average scores for the 9 reports using the pilot format were 24 percent lower than scores of the 15 traditional performance and accountability reports. Scores for reports in the traditional format were about the same in FY 2007 as in FY 2006. Average scores for reports using the pilot format fell by about 12 percent in FY 2007, compared to the scores on their traditional performance and accountability reports in FY 2006. The pilot format also forced the public to wait longer for performance information, since this information was released in February with the agencies’ congressional budget justifications, instead of in November as part of a performance and accountability report.

The principal benefit of the pilot format for the public that is evident at this point is its emphasis on the highlights document. We note, however, that the agencies producing traditional performance and accountability reports usually produced better highlights documents than the agencies using the pilot format. When done well, the highlights document is clearly a best practice that aids the public in understanding the agency’s performance information. We offer numerous suggestions for improving the highlights document.

A well-done highlights document improves disclosure to the public. The pilot format does not improve disclosure to the public. And good strategic planning is a crucial input into good performance reporting, because strategic planning establishes the goals and measures to be reported.

Those are the three biggest lessons from this year’s Performance Report Scorecard.

EVALUATION AND SCORING

The purpose of this assessment is not to evaluate or make judgments about the quality of actual results the agencies produced. Rather, our goal is simply to ascertain *how well the agencies' reports disclose to the public* the results they produced, so that policy makers and citizens may make informed judgments about the agencies' results. We review the reports solely from this perspective and not as accountants, government insiders, or experts on the functions of particular agencies.

TARGET AUDIENCE: THE GENERAL PUBLIC

A performance report necessarily addresses different audiences. Some are "stakeholders" with expertise in the agency's work, who seek an extensive level of understanding about the agency's performance and may be willing to plow through a lengthy, detailed, and technical report to get it. Others may be government insiders who know how to ferret out information that may be tucked away in congressional budget justifications or other esoteric documents.

From our perspective, though, the most important stakeholders are the ordinary citizens who pay the bills and deserve to know what each agency has accomplished in the last fiscal year. A report will not do well in our evaluation if it does not do a good job of informing the average citizen, even if it is informative for experts, insiders, or others who have more specialized knowledge. Of course, we do not expect tens of millions of fellow citizens to rush to agency Web pages to read these reports. Journalists, bloggers, and other writers can also play an important role in making agency results more widely accessible to the public. But like the general public, these readers are not agency insiders. The information should be accessible and understandable for those who wish to access it.

Reports that score high on our evaluation effectively communicate important performance results in a way that lay readers—ordinary citizens and taxpayers—can understand. This key trait is relevant to most categories in our *Scorecard*, and the best reports in this area tend to score well across the board. Reports that consistently score low do little to inform ordinary members of the public about important outcomes. Reports ranking in the middle may serve some audiences well but could do a better job of demonstrating the agency's value to ordinary citizens.

Specifically, in order to rank highly in this Scorecard, a report must:

- Use clear, concise presentation formats and language throughout that a layperson can follow and understand.
- Present a set of performance metrics that capture important public outcomes a lay reader can relate to and appreciate.
- Reinforce these performance metrics with clear narratives illustrating public benefits that flow from the agency's work.
- Enable the lay reader to readily grasp and assess progress toward outcomes.
- Provide confidence that the agency has adopted challenging measures, forthrightly acknowledges performance shortfalls, and takes steps to correct them.
- Provide confidence that the agency serves as a good steward of taxpayer resources by taking effective steps to resolve major management challenges.

Our research team used 12 evaluation factors grouped under three general categories of transparency, public benefits, and leadership:

1. Does the agency disclose its accomplishments in a transparent (easily understood) fashion?
2. Does the report focus on disclosing tangible public benefits (valued results) the agency produced?
3. Does the report show evidence of forward-looking leadership (guidance) that uses performance information to devise strategies for improvement?

TRANSPARENCY

Reports should be accessible, readable, and useable by a wide variety of audiences, including Congress, the administration, the public, news media, and other stakeholders. If a report fails to disclose significant achievements and problems to stakeholders, benefits or failures arising from agency activities will remain secret to all but a few insiders, and citizens will have no real opportunity to indicate their approval or disapproval.

PUBLIC BENEFITS

An agency's value to the public becomes clear only when its goals and measures are expressed in terms of the benefit produced or harm avoided for a particular set of clients or the public at large. To demonstrate openly how agency activities produce meaningful results for the community, reports should focus on "outcomes" (i.e., tangible benefits that matter in the lives of citizens) rather than on programs or activities as such. The reports should also clearly present the costs of achieving those results. These costs include both the agency's expenditures and other costs borne by the public at large. For regulatory agencies, much of this information should come from rigorous retrospective analysis of the actual effects of regulation, analogous to the information they are supposed to produce when conducting a Regulatory Impact Analysis prior to issuing a regulation. The ultimate objective of such reporting is to match outcomes with costs, so that policy makers and the public understand what citizens are paying to achieve various outcomes. Goals and measures that merely document agency activities, such as counts of checks processed or number of people enrolled in a program, assume that such activities automatically provide public benefits. Such an assumption can be incorrect for a wide variety of reasons. An agency report must highlight achievement of results; otherwise, it will not inform the public of the success or failure of its programs. Budget decisions that rely on such flawed information will fail to reflect realistic assessments of what agencies can accomplish with appropriations.

FORWARD-LOOKING LEADERSHIP

Agencies should use the performance information produced by their organizations to identify solutions to problems and to change future plans accordingly. The report should inspire confidence in an agency's ability to enhance citizens' quality of life commensurate with the resources they have entrusted to the agency. Among the factors that give such confidence is tangible evidence that the agency is using performance and financial data to improve management of its programs.

EVALUATING REPORTS UNDER THE 2007 PILOT FORMAT

For FY 2007, OMB allowed agencies to use an alternative reporting format under a pilot program.[7] Instead of producing a performance and accountability report that combines performance and financial information, agencies had the option of publishing detailed performance information along with their congressional budget justifications, producing a separate financial report, and producing a shorter "highlights" document intended to summarize both the performance and financial results. Nine agencies opted for the pilot approach. In addition, eight agencies that produced traditional performance and accountability reports also produced a highlights document, even though they were not required to do so. Both of these phenomena raised a novel issue: exactly what counts as a report for our evaluation purposes?

To answer this question, we began with the premise that our evaluation is intended to gauge the usefulness of these reports to the general public. The pilot format does not in any way absolve agencies of the responsibility to produce meaningful information for the public.

In Circular A-136, [8] OMB emphasized both that the pilot should do no harm by diminishing the public value of the performance reports, and that it is important to affirmatively enhance public value. Section II.1 of the Circular states:

> Using an alternative format should not in any way reduce the public's access to the detailed financial and performance information available in a consolidated PAR and should aim to make the presentation of performance information more meaningful. . . The goals of the pilot are to allow agencies to explore different formats to enhance the presentation of financial and performance information and make this information more meaningful and transparent to the public.[9]

Under OMB's guidance, the highlights document is to serve as the primary source of information for the general public. Section III of Circular A-136 states that the highlights document "should summarize key performance and financial information in a brief, user-friendly format that can be easily understood by a novice reader with little technical background . . . ." The Circular goes on to describe the minimum content of the highlights document:

- a clear, concise, and outcome-oriented explanation of the core public benefits that the agency seeks to achieve and a candid assessment of its progress;
- data and discussion of key performance measures, including trend data;
- a candid assessment of the agency's performance results against its goals;
- a summary of key management issues;
- references to internet links for more comprehensive information, including the full financial report, performance report, and performance and accountability report if an agency did one; and
- a summary of financial audit information.

For agencies participating in the pilot program, the highlights document is the primary document available and accessible to the public that explains performance and financial information. We recognized, however, that the highlights document might not contain as much detailed information as a traditional performance and

7. OMB's May 17, 2007 memorandum about the pilot program is available at http://www.whitehouse.gov/omb/memoranda/fy2007/par_alternative.pdf.

8. Available at http://www.whitehouse.gov/omb/circulars/a136/a136_revised_2007.pdf.

9. OMB Circular A-136, available at http://www.whitehouse.gov/omb/circulars/a136/a136_revised_2007.pdf, pp. 11-12.

accountability report. Therefore, for pilot agencies we started our evaluation with the highlights document. We then included information from other documents if the highlights document clearly indicated where and how the reader could find this information and presented it in an accessible and understandable form. The most common example of this was agency inspector general discussions of major management challenges. Pilot agencies included these in their financial reports, but usually informed the reader about this in the highlights document. Some also noted that more detailed performance information is available in a separate performance report. We did not consider information published in congressional budget justifications, because these documents are lengthy and difficult for non-specialists to navigate. None of the highlights documents specifically referenced or directly linked to a specific place in the budget justifications for the reader to find a particular type of information.

We treated the highlights document differently for agencies that produced traditional performance and accountability reports. For those agencies, the performance and accountability report should still be the primary means of communication with the public. After evaluating this report, we then examined the agency's highlights document to see if anything in that document merited a higher score. A highlights document might supplement a traditional performance and accountability report by improving readability. Several agencies did receive extra points on our readability criterion due to the quality of their highlights documents.

SCORECARD CRITERIA

TRANSPARENCY

1. Is the report easily accessible via the Internet and easily identified?
2. Is the report easy for a layperson to read and understand?
3. Are the performance data valid, verifiable, and timely?
4. Did the agency provide baseline and trend data to put its performance measures in context?

PUBLIC BENEFITS

5. Are the goals and objectives stated as outcomes?
6. Are the performance measures valid indicators of the agency's impact on its outcome goals?
7. Does the agency demonstrate that its actions have actually made a significant contribution toward its stated goals?
8. Did the agency link its goals and results to costs?

LEADERSHIP

9. Does the report show how the agency's results will make this country a better place to live?
10. Does the agency explain failures to achieve its goals?
11. Does the report adequately address major management challenges?
12. Does it describe changes in policies or procedures that will allow the agency to do better next year?

WHAT DID THE AGENCIES KNOW, AND WHEN DID THEY KNOW IT?

As in past years, the Mercatus Center notified federal agencies of the deadlines we would follow in evaluating the reports and the evaluation criteria we would employ.

For FY 2007, OMB required agencies (except those participating in the pilot for alternative performance reporting) to submit their reports to the president and Congress by November 15—approximately six weeks after the fiscal year ended. OMB required those agencies participating in the pilot program to release their "highlights" report (the primary document the Mercatus research team evaluated for agencies opting into the pilot) by February 1, 2008. Several agencies received extensions from OMB, permitting them to release their highlights on February 4.

On May 22, 2007, the Mercatus Center sent agencies a memorandum[10] describing how we expected to evaluate reports produced under the pilot format. The memo emphasized the key role of the highlights document as the general public's main source of performance information for the pilot agencies. For this reason, we advised that we would focus our reviews of pilot agency reports primarily on the highlights document. We also offered suggestions for the highlights document, which were consistent with OMB guidance.[11]

In September, the Mercatus Center notified each agency's chief financial officer via letter (and other individuals listed as agency GPRA contacts by e-mail) that the Mercatus research team would need either a paper or electronic copy of the report by December 1 for non-pilot agencies and February 1 for pilot agencies in order to include it in this year's evaluation. The letter also mentioned that traditional performance and accountability reports would need to be available on the web by December 15 and pilot reports by February 1 to earn credit on the first transparency criterion. The February 4 extension granted by OMB was recognized, on a case by case basis, for those agencies that requested it. All of the non-pilot agencies had their reports accessible on the web by December 15. Seven of the nine pilot agencies had their highlights reports accessible on the web by February 1 (or February 4 if OMB granted the agency an extension).

The September letter included an explanation of our evaluation criteria and noted that the quality of each year's reports "raises the bar" for subsequent years. The letter indicated how the Mercatus Center expected the FY 2007 evaluation process to differ from previous years, given OMB's pilot program. For the benefit of those agencies participating in OMB's pilot, the letter included a link to the May 22 memo detailing what the Mercatus research team would be looking for when reviewing highlights documents. Finally, the letter invited agency personnel to contact Mercatus Center staff with questions or comments about the criteria and the project, and many did so. Thus, agencies had ample notice about the criteria and deadlines.

10. The memo is available at http://www.mercatus.org/repository/docLib/20070523_Memo_to_agencies[1].pdf.

11. Suggestions for highlights documents are available at http://www.mercatus.org/repository/docLib/20070523_Suggestions_for_Highlights_Documents[1].pdf.

SCORING STANDARDS

Each report had the opportunity to earn up to 20 points in each of the three categories, for a maximum score of 60 points. Each category included four equally weighted evaluation factors, and scores of 1 through 5 (from inadequate to outstanding) were awarded on each evaluation factor. Thus, an agency could achieve a minimum score of 12 merely by producing a report.

THE 5-POINT SCALE

The 5-point rating scale for individual factors is intended to identify distinct levels of quality. The research team used the accompanying table to guide its scoring. A report that adequately meets all requirements would receive the middle score of 3 on each factor, resulting in a total score of 36. A score of 2 indicates that the report accomplishes some but not all of the objectives under a given criterion. A score of 1 indicates failure to provide much relevant information. A score of 4 indicates unusually good practices that are better than most, and a score of 5 indicates an especially superior presentation.

Even when a report receives a 5 on a particular criterion, that does not mean there is no room for improvement. A score of 5 indicates a potential best practice, but best practices should not be confused with perfection. We expect agency reporting practices to improve continually over time, and one of the goals of this Scorecard is to aid in the diffusion of best practices across agencies. Therefore, a practice that earned a 5 this year may only deserve a 4 or 3 in future years as it becomes standard for most agencies and new best practices emerge.

WEIGHTING THE EVALUATION FACTORS

To report the results of this study as transparently as possible, the researchers weighted the evaluation factors equally in calculating each agency's total score and rankings. Since the summary table reports scores for all three evaluation categories separately (transparency, public benefit, and forward-looking leadership), readers who believe that one factor is more important than others can apply whatever weights they wish to the separate scores and recalculate rankings accordingly.

WHAT DO THE SCORES MEAN?

5—OUTSTANDING

- Greatly exceeds expectations
- Opens a new field of information
- Sets a standard for best practice

4—VERY GOOD

- Exceeds expectations
- Has the potential to become a best practice
- Shows innovation and creativity
- Better than most

3—SATISFACTORY

- Meets expectations in all aspects
- Adequate, but does not exceed expectations

2—UNSATISFACTORY

- Fails to meet expectations
- May be adequate in some respects, but not all
- Produces partial information
- Does not fully disclose

1—INADEQUATE

- Fails to meet expectations
- Does not meet standards for adequate disclosure
- Shows no process or plans to overcome problems
- Omits critical information

In addition, in the interest of transparency, all reports were evaluated against a common scale, even though different agency missions may make it inherently more difficult to develop results-oriented goals and measures or collect appropriate data. For example, agencies that provide directly measurable services, such as the General Services Administration or Department of Veterans Affairs, might find it easier to identify and quantify their contributions than an agency like State or Homeland Security, whose results are less tangible. In reality, some agencies that seem to provide few services directly to members of the public have often produced highly-ranked reports, and some that arguably have a more direct effect on citizens' well-being have produced low-ranked reports.

INTERPRETING OUR FINDINGS

It is important to emphasize that our research team evaluated only *the quality of reporting, not the quality of results*. Therefore, it would be a mistake to conclude that the agencies with the highest-scoring reports necessarily produced the most or best results for the country.[12] A high scoring report reflects an agency's ability to explain its results in understandable and meaningful terms Americans can appreciate.

Similarly, it would also be inappropriate to draw policy conclusions from our analysis. We offer no recommendations on whether the federal government should or should not be engaged in its current menu of activities.

So what do the findings in this study really mean? By assessing the quality of agency reports, we are trying to evaluate whether the agencies are supplying the information that the public needs to understand what results the government is producing and what those results cost.

An additional word on information quality is also in order. Our researchers assessed the quality of each report's disclosure of data verification and validation procedures. However, in the interest of producing a timely study, we did not independently verify the performance information cited in each agency's report. The reports themselves should inspire confidence by indicating how data are verified and validated.

OUR CONSISTENCY CHECK

The Mercatus research team employed the same criteria to assess the FY 2007 agency reports that we used to evaluate prior year reports. However, generally in each succeeding year we have tightened our evaluation standards, for two reasons. One, the highest possible quality that can be achieved is unlimited, because creative innovators can always find ways to improve reporting practices and set new standards. Two, each year gives agencies an opportunity to learn from each others' best practices. If we did not continually raise our expectations, most reports could eventually receive mostly 5s. This Scorecard would then convey little information about the quality of different agencies' reports, and it would give little recognition or credit for those agencies that continue to raise the bar for quality reporting.

For these reasons, an agency had to improve the absolute quality of its FY 2007 report in order to receive the same numeric score it received for its FY 2006 report. That said, there were few new best practices in FY 2006, so the standards did not change a great deal for FY 2007. Nevertheless, if an agency receives a higher score, that score is a reliable indicator that the quality of its report has indeed improved.

12. Separately from our evaluation, several scholarly articles investigate whether there is any relationship the quality of performance reporting and agency performance. We discuss these on p.43 below.

Several factors help ensure that the scoring criteria are applied consistently from year to year. Two members of the Mercatus Center research team that evaluated the reports—Henry Wray and Valerie Richardson—have done so for the past five years. Patricia Kelley, the new research team member this year, has been on the advisory panel to the Scorecard project since 2002. The team cross-checked the 2007 evaluations against the previous year's in several ways. For each report, the research team generated an extensive set of notes documenting the reasons for each preliminary score on each criterion. The head of the research team reviewed this documentation for both the FY 2007 and FY 2006 reports to ensure that any scoring differences across years were justified by differences in the actual contents of the reports. The team discussed instances in which proposed scores differed substantially from the previous year's scores.

Finally, for each report, a member of our outside advisory panel with extensive experience in performance reporting reviewed the report, scoring, and documentation. Some scores were modified when the advisor reached different conclusions from the research team and offered persuasive reasons for the difference. Final scores thus reflect a careful review to ensure that the results of the scoring process are consistent with the goal of raising standards.