# INTRODUCTION

Outcome-oriented performance measurement isn't just a good idea; it's the law.

The Government Performance and Results Act (GPRA) requires federal agencies to produce strategic plans with performance measures, annual performance plans with performance goals, and annual performance reports that measure progress toward those goals. Measures are supposed to track the agency's "outputs, service levels, *and outcomes*."[1]

Ever since agencies issued their first performance reports for fiscal year 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, which accounted for virtually all federal outlays in fiscal year 2008.[2] The 24 comprise all cabinet departments plus the largest independent agencies.

Compared to 1999, we've come a long way. In fiscal year 1999, an agency could receive the highest possible score for accessibility of its report simply by posting the report on its Web page and making it easy to find. Today, that is a minimal requirement; best practices include making the report available on time, downloadable as a single document or individual sections, and providing clear information about whom to contact with comments or questions. Similarly, in fiscal year 1999, a report received the top score on linking results to costs because it broke down costs and personnel by program area. Today, the best reports break costs down by individual performance measure and present the information for multiple years.

To celebrate the tenth anniversary of GPRA performance reporting, our research team reevaluated the top four reports from fiscal year 1999 using fiscal year 2008's standards. The best report from fiscal year 1999, USAID, would have ranked only 16th in fiscal year 2008. Transportation and Veterans Affairs would have tied for 20th place, and Education would have finished 22nd. Based on this scoring experiment, we think it is safe to say that the average quality of performance reporting has increased by at least 75 percent.[3] Many reports have doubtless improved by much larger margins.

Despite this progress, outcome accountability for federal expenditures remains far short of ideal. In fiscal year 2008, 60 percent of federal expenditures (other than interest) were still covered by reports that received a score below the "satisfactory" level of 36, as figure 1 demonstrates.[4] This is a slight improvement from fiscal year 2007, when 65 percent of spending was covered by unsatisfactory reports. Only 7 percent of non-interest spending, or $207 billion, was covered by reports with a score of "very good"—48 or better.[5] This percentage is unchanged from last year.

1. GPRA sec. 1115. Emphasis added.
2. The principal parts of government not included in these 24 agencies are the judiciary, the legislature, the executive office of the president, and the independent agencies not among the 24 CFO Act agencies.
3. Details of the calculation can be found on 13.
4. A score of 3 on a criterion corresponds to a "satisfactory" rating. A report with an average score of 3 across all 12 criteria would earn a 36. Thus, a report must score 36 or higher to be classified as "satisfactory." For further explanation, see the appendix.
5. 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 the appendix.

Record budget deficits prompted by the financial crisis and recession make outcome accountability all the more imperative. One large part of the expanded deficit stems from the $787 billion worth of spending and tax reductions in the American Recovery and Reinvestment Act, passed in February 2009. The federal government has pledged "full transparency and accountability" in carrying out this legislation.[6] Guidance from the Office of Management and Budget (OMB) instructs agencies to account for the results of Recovery Act spending using their GPRA performance measures to the maximum extent possible.[7] Given this focus, it is worth considering how the quality of GPRA reporting matches up with Recovery Act spending and tax cuts.

Figure 2 classifies the Recovery Act's $334 billion in appropriations (listed in Division A of the legislation) according to the scores each agency received on this Scorecard for fiscal year 2008. Reports with scores in the "very good" range (48+ points) are most likely to achieve "full transparency and accountability." But as figure 2 shows, only 16 percent of the appropriations in the Recovery Act go to agencies whose reports met this standard in fiscal year 2008.

About 71 percent of appropriations go to agencies whose reports received a "satisfactory" score of 36 or better in fiscal year 2008. This looks like a significant improvement compared to the percentage reported in a March 2009 *Mercatus on Policy* brief, which found that 74 percent of spending went to agencies whose reports scored *below* satisfactory in the fiscal year 2007 Scorecard.[8] Most of the difference occurs because two agencies receiving large amounts of stimulus money improved their scores by several points in fiscal year 2008. Education's score rose from 32 to 37, and Energy's rose from 31 to 36.

The Recovery Act also makes direct grants to states and expands federal payments to individuals (in Division B of the legislation). It is not clear if the results of these expenditures, totaling approximately $270 billion, will also be tracked using GPRA outcome measures. Another $212 billion of the Recovery Act is tax provisions, which are still not covered by GPRA performance measures 16 years after passage of the law.[9]

For at least 787 billion new reasons, transparent reporting of outcomes is now more important than ever.

6. See http://www.recovery.gov.

7. Office of Management and Budget, "Memorandum for the Heads of Departments and Agencies," February 18, 2009, http://www.recovery.gov/files/Initial%20Recovery%20Act%20Implementing%20Guidance.pdf.

8. Christina Forsberg and Stefanie Haeffele-Balch, "Accountability and Transparency in the American Recovery and Reinvestment Act," *Mercatus on Policy* 38 (March 2009), 3, http://www.mercatus.org/PublicationDetails.aspx?id=26482.

9. The 2009 Analytical Perspectives volume of the Budget of the United States Government explores some broad suggestions for measuring the outcomes of tax expenditures under GPRA and promises to keep working on it: "The above illustrative discussion, although broad, is nevertheless incomplete, omitting important details both for the provisions mentioned and the many that are not explicitly cited. Developing a framework that is sufficiently comprehensive, accurate, and flexible to reflect the objectives and effects of the wide range of tax expenditures will be a significant challenge. OMB, Treasury, and other agencies will work together, as appropriate, to address this challenge." *Budget of the United States Government*, *Fiscal Year 2009*, Analytical Perspectives, 328.

**Figure 1: Fiscal Year 2008 Spending ($Billions) and Quality of Disclosure**

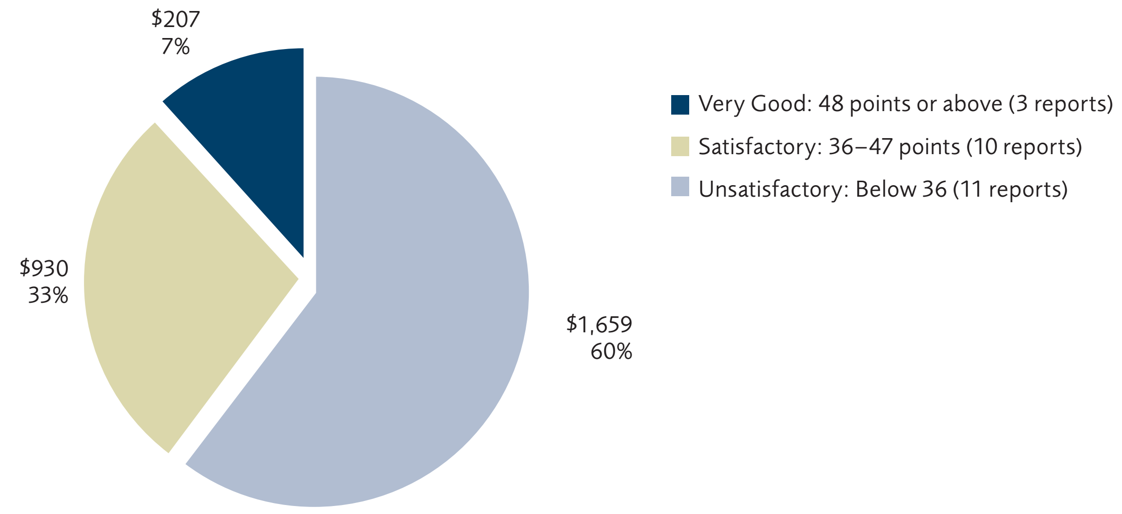
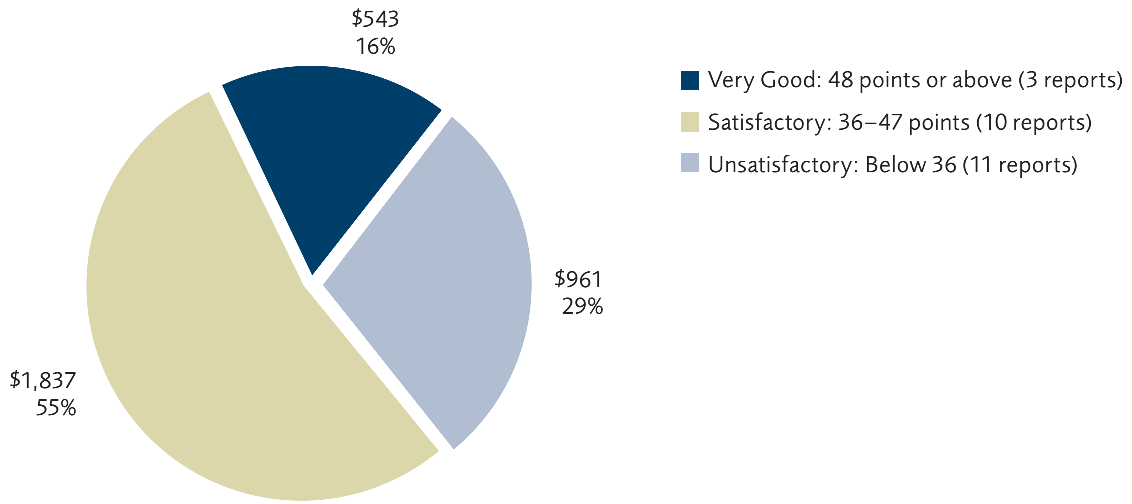



**Figure 2: Recovery Act Appropriations ($Billions) and Quality of Disclosure**



# EVALUATION AND SCORING

FOR THE PAST ten years, the Mercatus Center's Performance Report Scorecard has evaluated how well performance reports disclose to the public the outcomes the agencies produce. The Scorecard does not offer an opinion on the quality of agencies' performance, nor does it express views on what activities the government should or should not undertake. We assess the quality of disclosure, not the quality of results.

Reports are evaluated according to 12 criteria grouped into three categories:

1. Transparency: How easily can a non-specialist find and understand the report?
2. Public benefits: How well does the report document the outcomes the agency produces for the public and compare them with costs?
3. Leadership: How well does the report demonstrate that agency managers use performance information to make decisions?

On each of the 12 criteria, a report can achieve a score ranging from 1 point (no useful content) to 5 points (potential best practice). Thus, total report scores can range from 12 to 60 points. Over time, the research team tightens its scoring standards to reflect the development of new best practices.

A more detailed description of the scoring criteria and evaluation methods is available in previous editions of the Scorecard and is reproduced as an appendix this year.

### Scorecard Criteria

TRANSPARENCY

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

PUBLIC BENEFITS

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

LEADERSHIP

9. Vision: Does the report show how the agency's results will make this country a better place to live?
10. Explain Failures: Does the agency explain failures to achieve its goals?
11. Major Management Challenges: Does the report adequately address major management challenges?
12. Improvement Plans: Does the report describe changes in policies or procedures to do better next year?

# SCORING SUMMARY

## TOP THREE SWAP PLACES

The top three reports—from Labor, Veterans Affairs, and Transportation—continue to outdistance the others by a large margin. But once again this year, the top three switched positions. Labor's report took first place with a score of 56, the highest score ever awarded in the ten-year history of the Scorecard. Veterans Affairs took second with a score of 54, and Transportation earned third place with a score of 53. Thus, just three points separate these reports. Homeland Security and the NRC tied for fourth place with 40 points—a modest improvement over their scores for fiscal year 2007.

## MOST-IMPROVED REPORTS

The biggest gainers among agencies that made meaningful improvements to their reports were State (+6 points), Energy (+5 points), and Interior (+5 points). Education and USAID each gained 4 points. Notably, three of these used the pilot reporting format that OMB first introduced in fiscal year 2007 instead of producing a traditional performance and accountability report (PAR). There is no striking pattern explaining the score increases for the pilot agencies; greater access to information from source documents (particularly performance reports) probably is the main factor. Education's gain resulted primarily from a new strategic plan, which significantly enhanced the department's performance metrics.

Defense was the biggest individual gainer, increasing its score 9 points from 17 last year to 26 this year. Nevertheless, Defense still produced one of the worst reports. The department's score increased from 1 to 5 on criterion 1 (accessibility) by posting its citizens' report on time and satisfying the other evaluation factors for that criterion. The remaining gains were increases from 1 (no relevant content) to 2 (minimal content) in five other categories. This report still has far to go before it regains the ground it has lost since fiscal year 2006 when it earned a score of 32.

There were no dramatic individual score decreases. HUD lost the most ground (-4 points), dropping in four categories. Several agencies dropped 1 or 2 points.

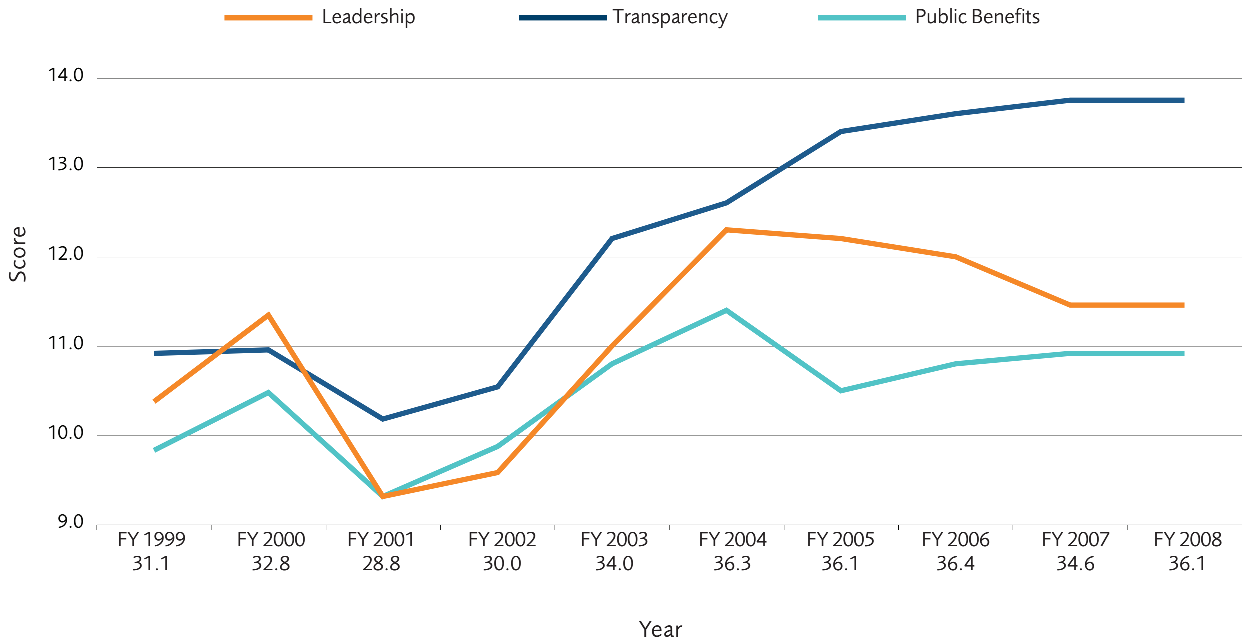
## CHURN IN THE RANKINGS

Below the top five reports, fiscal year 2008 saw significant churn in the rankings. This occurred because one group of agencies made measurable improvements while others scored about the same as in fiscal year 2007. State, Energy, Interior, and Education leapt by 8 to 12 places in the rankings, and USAID rose by 4 places. Because these reports improved their scores, the others plummeted by between 4 and 9 spots in the rankings.

## AVERAGE SCORES RISE

The average score for all 24 agencies was 36.1. As figure 3 shows, this is an increase from last year's average (34.6) and a slight decrease from fiscal year 2006 (36.4). The increase from fiscal year 2007 is attributable primarily to improvement in the pilot agency scores. The average score for traditional performance and accountability reports increased 1 point, from 37.3 last year to 38.3 this year. The average pilot agency score increased 2.6 points, from 30.0 last year to 32.6 this year.

**Figure 3: A Year of Recovery**




## Report quality clusters in the middle

Almost half of the report scores cluster around the satisfactory score of 36. Four reports earned 37 points, four earned 36 points, one earned 35 points, and one earned 34 points. As figure 4 shows, this led to a noticeable increase in the number of reports scoring satisfactory or better in fiscal year 2008—reversing the past several years' trend. About half of the reports (13) achieved a satisfactory score for fiscal year 2008, up from just 7 reports in fiscal year 2007. Ten reports scored between 24 and 35 points, down from 15 in fiscal year 2007. Only one report scored below 24 in fiscal year 2008.

In general, transparency has improved at a faster rate and achieved a higher score than public benefits or forward-looking leadership. This probably indicates that it is easier to improve on factors like availability and readability than on substantive matters like definition of outcomes or analysis of performance results. Indeed, even many of the low-ranking agencies earned scores of 4 or 5 on some of the transparency criteria.

## Opportunities for progress

There is still substantial room for progress. Table 4 reveals that average scores for eight criteria are below 3.

For most criteria, average scores increased in fiscal year 2008. Thanks largely to these improvements, below-satisfactory scores on many criteria are nevertheless close to 3. These include verification and validation (2.92 points), baseline and trend data (2.88 points), outcome-oriented measures (2.88 points), explanation of management challenges (2.92 points), and improvement plans (2.83 points).

**Figure 4: More Reports Are Satisfactory**

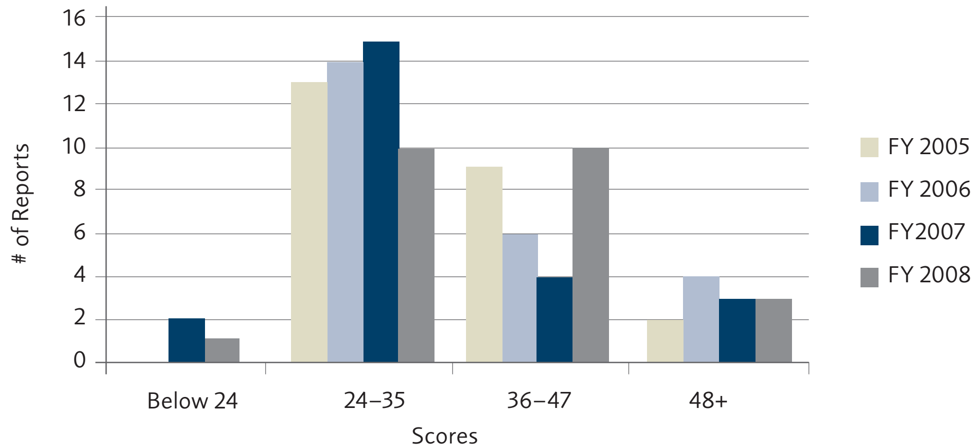



**Table 4: Eight Criteria Below Satisfactory**

| CRITERION | FISCAL YEAR 2007 | FISCAL YEAR 2008 | CHANGE 2007–08 | % CHANGE 2007–08 |
|---|---|---|---|---|
| 1. Accessibility | 4.08 | 4.38 | 0.29 | 7 |
| 2. Readability | 3.63 | 3.58 | -0.04 | -1 |
| 3. Verification & Validation | 2.46 | 2.92 | 0.46 | 19 |
| 4. Baseline/Trend Data | 2.67 | 2.88 | 0.21 | 8 |
| 5. Outcome Goals | 3.13 | 3.21 | 0.08 | 3 |
| 6. Outcome Measures | 2.67 | 2.88 | 0.21 | 8 |
| 7. Agency-Affected Outcomes | 2.67 | 2.63 | -0.04 | -2 |
| 8. Linkage to Costs | 2.13 | 2.21 | 0.08 | 4 |
| 9. Vision | 3.00 | 3.08 | 0.08 | 3 |
| 10. Explain Failures | 2.67 | 2.63 | -0.04 | -2 |
| 11. Management Challenges | 2.83 | 2.92 | 0.08 | 3 |
| 12. Improvement Plans | 2.67 | 2.83 | 0.17 | 6 |

*Scores are rounded*

## YOU'VE COME A LONG WAY, BABY!

Federal performance reporting has improved substantially since fiscal year 1999—much more than the 15-percentage-point increase in average score over the past ten years would indicate. Table 5 shows the change in each report's score between fiscal year 1999 and fiscal year 2008. Large drops in scores for three reports—USAID, Defense, and SBA—help pull down the average score increase. For the 17 reports whose scores improved, the average increase was 8.94 points, almost double the average increase of 4.84 points for all 24 reports. Nine reports achieved double-digit increases in their scores.

All of these score data understate the full extent of improvement because the research team tightens the scoring criteria over time as new best practices emerge. To gain a better understanding of how the quality of disclosure has improved, we performed both a quantitative and a qualitative evaluation.

**Table 5**

## Fiscal Year 2008 Scores & Rankings: Comparison to Fiscal Year 1999

Highest Rank = 1; Lowest = 24. Maximum Possible Score = 60; Minimum = 12.

| | Fiscal Year 2008 | | Fiscal Year 1999 | | | |
|---|---|---|---|---|---|---|
| | TOTAL SCORE | RANK | TOTAL SCORE | RANK | CHANGE IN SCORE | CHANGE IN RANKING |
| Labor | 56 | 1 | 36 | 5 | +20 | +4 |
| Veterans Affairs | 54 | 2 | 48 | 3 | +6 | +1 |
| Transportation | 53 | 3 | 51 | 2 | +2 | -1 |
| Homeland Security* | 40 | 4 | 27 | 22 | +13 | +18 |
| NRC | 40 | 4 | 25 | 17 | +15 | +13 |
| Education | 37 | 6 | 37 | 4 | 0 | -2 |
| Interior | 37 | 6 | 31 | 11 | +6 | +5 |
| State | 37 | 6 | 25 | 17 | +12 | +11 |
| Treasury | 37 | 6 | 36 | 5 | +1 | -1 |
| Energy | 36 | 10 | 27 | 14 | +9 | +4 |
| EPA | 36 | 10 | 31 | 11 | +5 | +1 |
| Health & Human Services | 36 | 10 | 24 | 20 | +12 | +10 |
| USAID | 36 | 10 | 52 | 1 | -16 | -9 |
| Commerce | 35 | 14 | 22 | 22 | +13 | +8 |
| Justice | 34 | 15 | 23 | 21 | +11 | +6 |
| Agriculture | 33 | 16 | 22 | 22 | +11 | +6 |
| GSA | 32 | 17 | 32 | 9 | 0 | -8 |
| NSF | 32 | 17 | 21 | 24 | +11 | +7 |
| Social Security | 32 | 17 | 33 | 8 | -1 | -9 |
| NASA | 31 | 20 | 27 | 14 | +4 | -6 |
| OPM | 28 | 21 | 27 | 14 | +1 | -7 |
| HUD | 27 | 22 | 28 | 13 | -1 | -9 |
| Defense | 26 | 23 | 34 | 7 | -8 | -16 |
| SBA | 22 | 24 | 32 | 9 | -10 | -15 |
| Average | 36.13 | | 31.29 | | 4.83 | |
| Median | 36.00 | | 29.50 | | 6.50 | |

**Since DHS did not exist in 1999, the chart shows its score and rank from fiscal year 2003, the first year its report was included in the Scorecard.*

## BY THE NUMBERS

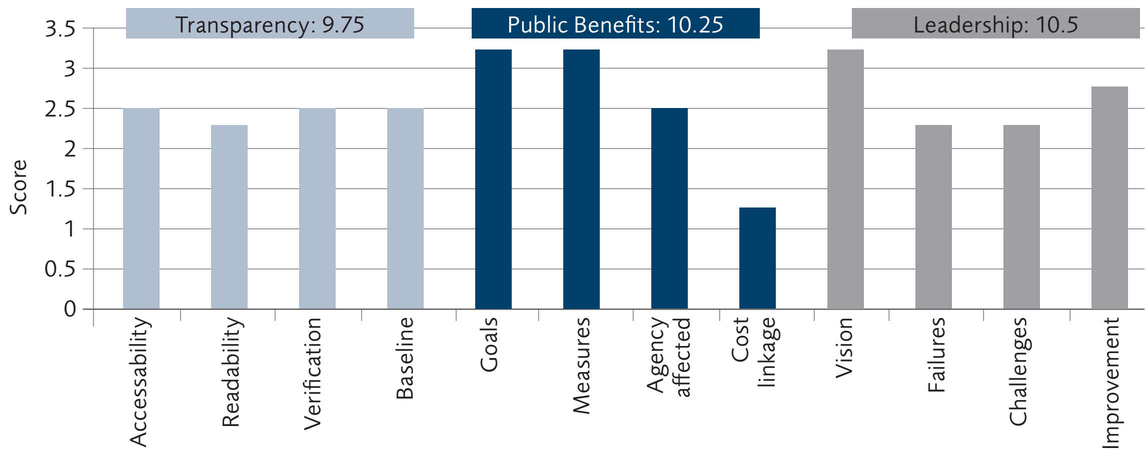
For its quantitative evaluation, the research team re-examined the top four reports from fiscal year 1999 using the same standards that applied to the fiscal year 2008 reports.[10] Comparing the scores the same report receives under two different years' standards should reveal how much the standards have tightened over ten years. Comparing reports the same agency produced in two different years, both evaluated under the fiscal year 2008 standards, should reveal how much that agency's report actually improved.

Table 6 shows the results. Evaluated by fiscal year 2008 standards, the best fiscal year 1999 report (from USAID) would have ranked 16th in fiscal year 2008, with just 33 points out of a possible 60. Transportation and Veterans Affairs received 51 and 48 points respectively in fiscal year 1999, ranking second and third. Evaluated under fiscal year 2008 standards, they both received 31 points and would have tied for 20th place in fiscal year 2008. Education, the fourth-place report in fiscal year 1999, received 37 points. Evaluated by fiscal year 2008 standards, it received 27 points and would have ranked 22nd in fiscal year 2008.

We can measure how these specific agencies' reports have improved by comparing the scores for their fiscal year 2008 reports with the scores on their fiscal year 1999 reports evaluated on the same 12 criteria under fiscal year 2008 standards. Using the scores reported in table 6, USAID's report improved by about 9 percent over ten years (from 33 to 36 points), Education's report improved by 37 percent (from 27 to 37 points), Transportation's report improved by 71 percent (from 31 to 53 points), and Veterans Affairs' report improved by 74 percent (from 31 to 54 points).

Figure 5 shows how these four reports scored on individual criteria. Note that average scores on criteria 5 (outcome-oriented goals), 6 (outcome-oriented measures), and 9 (vision) all exceed the satisfactory score of 3—even when evaluated by fiscal year 2008 standards. This suggests that in the early days of GPRA, the higher-ranking agencies got off to a good start in formulating outcome-oriented goals and measures. Also, the average score on criterion 8 (linkage of results to costs) barely exceeds 1. This indicates that, compared to current practice, even the top scorers had little cost-related content in fiscal year 1999. Third, the readability score is also quite low—and the research team agreed that readability has improved substantially since fiscal year 1999.

**Figure 5: Average Scores for Re-evaluated FY 1999 Reports**




10. This was considered easier than transporting some fiscal year 2008 reports back in time so the original research team could evaluate them in 1999.

**Table 6**

## Top Four 1999 Reports Would Rank Low Today

Highest Rank = 1; Lowest = 24. Maximum Possible Score = 60; Minimum = 12

| | Fiscal Year 2008 | | | | |
|---|---|---|---|---|---|
| | **Transparency** | **Public Benefits** | **Leadership** | **Total** | **Rank** |
| Labor | 20 | 19 | 17 | 56 | 1 |
| Veterans Affairs | 19 | 16 | 19 | 54 | 2 |
| Transportation | 16 | 20 | 17 | 53 | 3 |
| Homeland Security | 15 | 13 | 12 | 40 | 4 |
| NRC | 15 | 13 | 12 | 40 | 4 |
| Education | 14 | 12 | 11 | 37 | 6 |
| Interior | 16 | 10 | 11 | 37 | 6 |
| State | 15 | 10 | 12 | 37 | 6 |
| Treasury | 14 | 10 | 13 | 37 | 6 |
| Energy | 13 | 11 | 12 | 36 | 10 |
| EPA | 13 | 11 | 12 | 36 | 10 |
| Health & Human Services | 13 | 13 | 10 | 36 | 10 |
| USAID | 15 | 10 | 11 | 36 | 10 |
| Commerce | 15 | 10 | 10 | 35 | 14 |
| Justice | 15 | 8 | 11 | 34 | 15 |
| Agriculture | 12 | 10 | 11 | 33 | 16 |
| USAID 1999 | 9 | 11 | 13 | 33 | 16 |
| GSA | 11 | 12 | 9 | 32 | 17 |
| NSF | 15 | 7 | 10 | 32 | 17 |
| Social Security | 12 | 8 | 12 | 32 | 17 |
| NASA | 11 | 8 | 12 | 31 | 20 |
| Transportation 1999 | 9 | 12 | 10 | 31 | 20 |
| Veterans Affairs 1999 | 11 | 10 | 10 | 31 | 20 |
| OPM | 11 | 8 | 9 | 28 | 21 |
| HUD | 11 | 8 | 8 | 27 | 22 |
| Education 1999 | 10 | 8 | 9 | 27 | 22 |
| Defense | 11 | 7 | 8 | 26 | 23 |
| SBA | 8 | 8 | 6 | 22 | 24 |
| Fiscal Year 2008 Average | 13.8 | 10.9 | 11.5 | 36.1 | |
| Fiscal Year 2008 Median | 14.0 | 10.0 | 11.0 | 36.0 | |

On average, we estimate that report quality may have improved by about 75 percent. Scores for the four re-evaluated fiscal year 1999 reports averaged one-third lower under the fiscal year 2008 standards than under the fiscal year 1999 standards. If we assume that using the fiscal year 2008 scoring standards would have reduced all fiscal year 1999 scores by one-third, the average fiscal year 1999 score using fiscal year 2008 standards would have been 20.65 instead of the 31.29 shown in table 5. An increase from 20.65 to the average fiscal year 2008 score of 36.13 implies that the average quality of performance reports improved by at least 75 percent.[11] Of course, that means the quality of some agencies' reports increased by even more than that amount, others by less.

## Qualitative evaluation

Since the Mercatus Scorecard explicitly identifies best practices every year, it is not difficult to compare some fiscal year 1999 best practices with their fiscal year 2008 counterparts. This comparison reveals how the "state of the art" in the best reports has advanced during the past decade. Except for criterion 1 (accessibility), the best practices in the list below are not usually typical practices. In most cases, only a few reports in each year used the best practices. The list below provides a summary; more detailed descriptions can be found in the fiscal year 1999 Scorecard (http://www.mercatus.org/PublicationDetails.aspx?id=21144) and the "Strongest and Weakest Scores" section of this report.

### Criterion 1: Accessibility

Fiscal Year 1999—Labor and OPM had their reports available online with obvious links from their home pages. Sixteen agencies made their reports available online, but only four earned a score of 5 for clearly labeling the report as the annual performance report and making it easy to find.

Fiscal Year 2008—Thirteen agencies earned a score of 5 because they posted their reports on time, created a direct link on the home page, permitted downloads as both single and multiple files, and provided contact information for questions or comments.

### Criterion 2: Readability

Fiscal Year 1999—USAID produced a lengthy report with a massive amount of detail, but the organization, headings, sidebars, tables, and charts made the report easy to scan and digest. Relative lack of jargon made the report understandable to readers who are not specialists in foreign aid.

Fiscal Year 2008—Justice and Labor are the only reports to score a 5. The Management's Discussion and Analysis section and the Performance section of Justice's report are much shorter than in most PARs. Justice's report includes 25 key measures and presents goals, measures, and results in convenient tables. Labor's report is visually appealing with helpful tables and graphics. The secretary's letter describes key achievements over time as well as fiscal year 2008 results for each strategic goal. Both reports are complemented by excellent citizens' reports that include the two-page snapshot, concise summaries of results, and reader-friendly links to additional information.

11. If the average fiscal year 1999 score using fiscal year 2008 scoring standards would have been 20.65, the percentage improvement is calculated by subtracting this score from the actual fiscal year 2008 score (36.13), then dividing this difference by 20.65.

### Criterion 3: Verification and validation

Fiscal Year 1999—Education's report presented a thorough discussion of validation procedures and data limitations. The report acknowledged data deficiencies and outlined plans to develop new databases, improve data quality, and ensure the accuracy of financial statements.

Fiscal Year 2008—Labor's report assesses the quality of its data for each performance goal and provides data sources for each measure. Tables in the Veterans Affairs report provide data definitions, verification and validation information, and data limitations.

### Criterion 4: Baseline and trend data

Fiscal Year 1999—Transportation's report presented trend data for as far back as five or ten years. Defense juxtaposed three years of actual data with the goals for each program for the next three years. Veterans Affairs, Education, and USAID also provided useful baseline data.

Fiscal Year 2008—Labor and Veterans Affairs both present four years' worth of baseline data for key measures. Data include targets, actual results, and (in Labor's case) costs. Labor and Treasury use graphics to show performance trends. Veterans Affairs shows long-range targets for each measure, and Transportation presents a fiscal year 2009 forecast for each measure.

### Criterion 5: Outcome-oriented goals

Fiscal Year 1999—All five of Transportation's strategic goals were outcome oriented. Most of Veterans Affairs' and USAID's were, too.

Fiscal Year 2008—The vast majority of Labor's and Transportation's goals and objectives are clear, measurable outcomes that the general public can appreciate and understand, such as safer workplaces, employment of people who were in job training programs, reduced highway congestion, and reduced transportation accidents and fatalities.

### Criterion 6: Outcome measures

Fiscal Year 1999—Most measures in the Veterans Affairs and Transportation reports were outcomes or closely related to outcomes.

Fiscal Year 2008—Labor's measures for three of its four strategic goals are outcomes; the fourth has a mix of outcome, activity, output, and efficiency measures. Transportation's measures for three of five strategic goals are all outcomes; for two other goals, about half the measures are end or intermediate outcomes.

### Criterion 7: Agency-affected outcomes

Fiscal Year 1999—USAID explained how its activities contributed to results in specific cases, and also acknowledged other factors that may have contributed to the results.

Fiscal Year 2008—Four agencies consistently describe how their activities led to observed results on highly outcome-oriented performance metrics.

### Criterion 8: Results linked to costs

Fiscal Year 1999—The Veterans Affairs report included charts that juxtaposed personnel and costs by program area with performance measures for each program area.

Fiscal Year 2008—Labor and Transportation both assign costs to strategic goals and most individual performance measures. Labor's report contains this information for several years.

### Criterion 9: Vision

Fiscal Year 1999—Transportation made a case for how it affects Americans' quality of life by writing each strategic goal so that it identifies a broad result of interest to citizens and then explains the means by which the department hopes to accomplish the result.

Fiscal Year 2008—Narratives in both the Transportation and Veterans Affairs reports cite specific, major accomplishments that affect citizens' quality of life. Narratives and vignettes are linked to outcome-oriented performance metrics, and performance metrics demonstrate that the narratives describe general results rather than isolated, anecdotal successes.

### Criterion 10: Explanation of failures

Fiscal Year 1999—USAID's report provided explanations when targets were not met and described planned remedies for unmet targets.

Fiscal Year 2008—Veterans Affairs' report includes a "Performance Shortfall Analysis" that describes reasons for shortfalls and outlines steps and a timeline to remedy them. The report even offers performance improvement steps in cases where targets were met. NASA, Transportation, and Treasury also identify shortfalls on performance measures, explain the shortfalls, and outline plans to improve.

### Criterion 11: Major management challenges

Fiscal Year 1999—Transportation, USAID, and Veterans Affairs all include thorough discussions of management challenges identified by the agency inspector general and the GAO.

Fiscal Year 2008—Best practices include an insightful inspector general report that lists major management challenges and assesses the agency's progress toward resolving them, an agency self-assessment of progress, a description of actions taken, a timeline for resolving each challenge, and an explanation of how each challenge is related to the agency's strategic goals. No agency did all of these things, but various combinations can be found in the reports from Labor, EPA, Transportation, and Veterans Affairs. The best example of an insightful inspector general report is the one for SBA. Unfortunately, the SBA citizens' report does not reference this very useful document, which appears in the agency's financial report.

### Criterion 12: Improvement plans

Fiscal Year 1999—Discussion of each measure in Transportation's report included fiscal year 1999 results, projected performance in fiscal year 2000, and actions planned to achieve goals in fiscal year 2001.

Fiscal Year 2008—Reports from Labor, Veterans Affairs, and Transportation present improvement strategies for performance shortfalls and major management challenges. They also describe broader challenges the agencies expect to face in the future and offer plans for addressing them.

This qualitative description of best practices is consistent with the quantitative score assessment: Performance reporting has made significant progress since fiscal year 1999.