Introduction: If Seven's So Lucky, Then Why Aren't We There Yet?

The past year has seen a dramatic change in public awareness of government accountability. Lobbying scandals, bridges to nowhere, disaster relief disasters, and earmarks that take precedence over reconstruction have led both the average citizen and the media to ask, "Is anybody minding the store?"

Perhaps surprisingly, the answer is yes. Ever since Congress passed the Government Performance and Results Act in 1993, career civil servants and political appointees alike have labored 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 the public. And 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 (CFO Act), which accounted for 98.8 percent of federal outlays in fiscal 2005.[1]

The quality of disclosure in these reports has improved, albeit not as rapidly as we had hoped. Comparing the quality of agency performance and accountability reports with federal expenditures provides a rough idea of how much federal spending is devoted to programs with well-documented outcomes. In fiscal 2005, after seven years of reporting, only 15 percent of non-interest federal spending by the 24 agencies was covered by reports with total scores of "satisfactory" or better in our evaluation.[2] Though covering less than one-quarter of non-interest expenditures, these reports came from 11 of the 24 Chief Financial Officers Act agencies, as Figure 1 shows. Yet only a tiny fraction of non-interest spending—three percent—was covered by the two reports whose scores put them in the highest category—"very good."[3] (See Figure 2.) A person relying on the reports for performance information would have to conclude that sub-standard disclosure underlies huge portions of federal spending.

[1] Figure calculated from data in *Budget of the United States Government, Fiscal Year 2007*, Historical Tables, Table 4.1. Denominator is outlays plus undistributed offsetting receipts.

[2] 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 p. 7 .

[3] 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 p. 7 below.

Figure 1: Fiscal 2005 spending ($billions) covered by "satisfactory" disclosure (36 points or above out of 60)

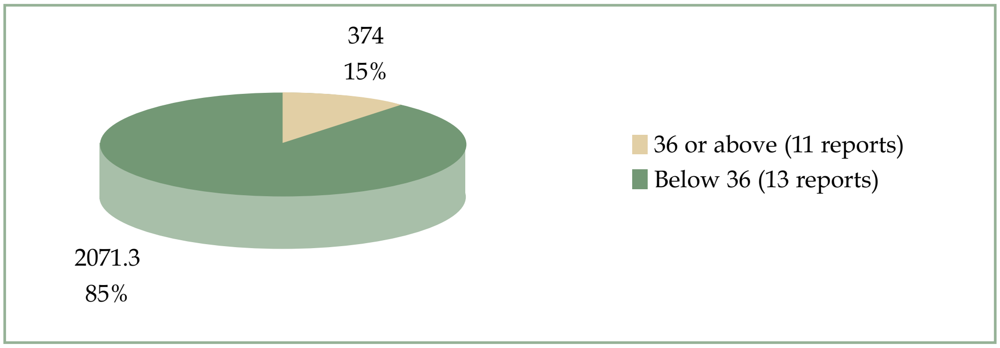
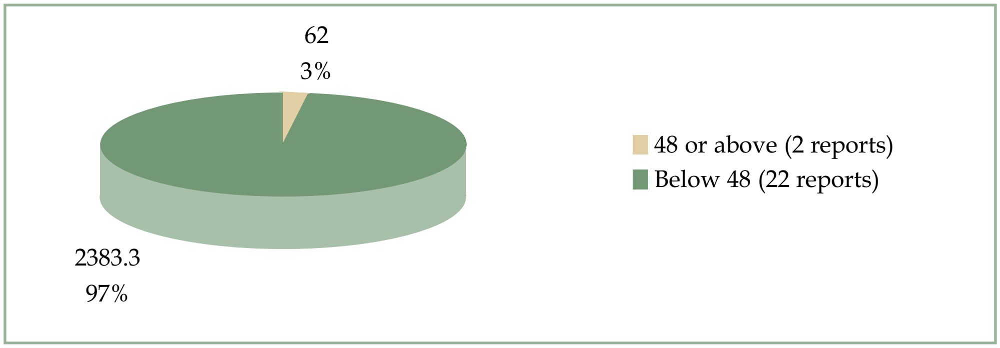


Figure 2: Fiscal 2005 spending ($billions) covered by "very good" disclosure (48 points or above out of 60)



Our approach emphasizes the importance of measuring results rather than intentions. The part of our evaluation that focuses on the public benefit criteria is the part that most directly evaluates the extent to which the reports focus on outcomes. We assess whether the reports articulate outcome-based goals and measures, together with linkage of costs to outcomes. Figures 3 and 4 show how much of the 24 agencies' expenditures are covered by reports that score well on these criteria, which focus on articulation of outcomes and connection of costs to outcomes. In fiscal 2005, only eight percent of non-interest federal spending by the 24 agencies occurred in the six agencies that produced reports earning a satisfactory or better score on our public benefits reporting criteria. Just three percent of non-interest spending was covered by reports earning a score of "very good" or better.

Figure 3: Fiscal 2005 spending ($billions)
covered by reports with "satisfactory" public benefits score (12 or above out of 20)

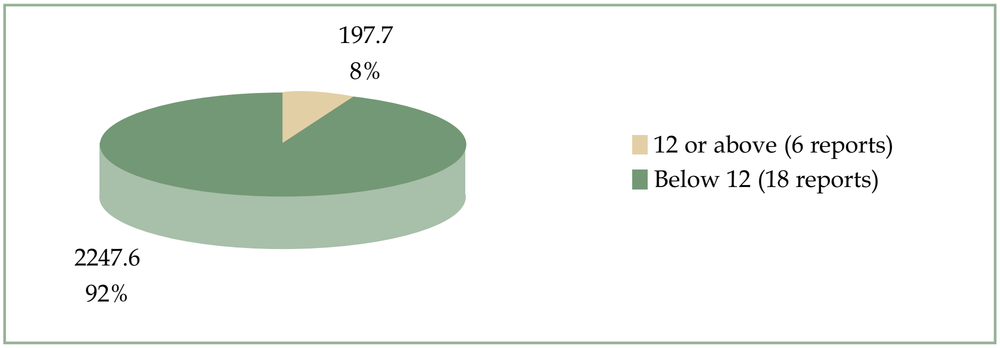
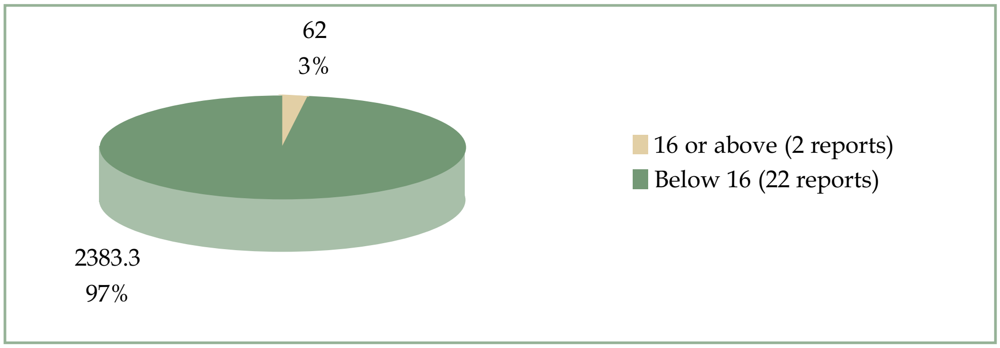


Figure 4: Fiscal 2005 spending ($billions)
covered by reports with "very good" public benefits score (16 or above out of 20)



Given the paucity of links between outcomes and costs in most of the reports, it's tempting to conclude that vast swaths of federal spending are essentially "faith-based" initiatives. Intentions and values, rather than systematic proof of actual outcomes, drive much of the support for these programs.

Unfortunately, intentions do not equal results. A misguided idealist can jump off a cliff with the intention of flying to Mars by flapping his arms, but wishing does not make it so—and neither does the nobility of his intentions or values. Whether the goal is to fly to Mars, house the homeless, or reduce the threat of terrorism, no substitute exists for real information that shows whether the means we've selected have actually accomplished our ends.

An old proverb attributed to several great thinkers warns, "The road to hell is paved with good intentions."[4] And because we have made policy decisions based on intentions rather than results, now there's hell to pay. But recent events, and the failures associated with them, create an excellent opportunity to refocus policymakers' attention on outcomes and accountability. After all, what better way to promote accountability and restore confidence in government than with real facts about the actual results and costs of federal initiatives?

Effective accountability requires that agencies present full, concise, relevant, accurate, and reliable disclosure of the benefits created for the public, as well as the costs of producing those benefits. In some sense, OMB's PART initiative has outdone agencies' Performance and Accountability Reports. OMB also seems to be ahead of Congress in using performance information to make budget decisions. PART analyses line up information on program design, spending, and performance in a way that few agency reports have matched. It's now a horse race to see whether PART or PARs will be the first to link dollars directly with outcomes. Without such information, it is unlikely that policymakers could allocate federal resources in ways that advance government's contribution to citizens' quality of life.

[4] Contrary to popular belief, the origin of the proverb is neither Karl Marx nor Samuel Johnson. See http://www.samueljohnson.com/road.html and http://www.everything2.com/index.pl?node_id=747936.

Evaluation Criteria

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 policymakers and citizens may make informed judgments about the agencies' results. Our research team utilized 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. The ultimate objective of such reporting is to match outcomes with costs, so that policymakers 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 these programs.

Transparency

1. Is the report easily accessible via the Internet and easily identified as the agency's Annual Performance and Accountability Report?
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 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 fiscal 2005, the Office of Management and Budget required agencies to submit their reports to the president and Congress by November 15—approximately six weeks after the fiscal year ended. 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 a copy of the report by December 1 in order to include it in this year's evaluation. The letter also mentioned that reports would need to be available on the web by December 15 to earn credit on the first transparency criterion. A followup reminder went out via e-mail on November 9. All agencies had their reports accessible on the web by December 15.

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. It also thanked agencies for their continued participation in our study and reiterated our goal of continually improving the quality of reporting. Finally, the letter invited agency personnel to contact Mercatus Center staff with questions or comments about the criteria, and many did so. Thus, agencies had ample notice about the criteria and deadlines.

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 poor to excellent) 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 2 indicates that the report accomplishes some but not all of the objectives under a given criterion. A 1 indicates failure to provide much relevant information. A 4 indicates unusually good practices that are better than most, and a 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 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.

What do the scores mean?

5

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

4

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

3

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

2

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

1

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

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.

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 direct 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 the State Department, dealing in more intangible results In reality, some agencies that seem to provide few services directly to members of the public, such as State, have eventually produced highly-ranked reports, and some that arguably have a more direct effect on citizens' well-being, such as the Social Security Administration, 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 result*s. 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. Ideally, an agency's report reflects more about its managers' capabilities than just their ability to write reports. A high scoring report reflects an agency's ability to translate what it does into understandable and meaningful results that 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 the agencies that are supplying the information that Congress and the public need to make informed funding, budgeting, and policy decisions.

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.

How valid are *our* data?

The Mercatus research team employed the same criteria to assess the fiscal 2005 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 fiscal 2005 report in order to receive the same numeric score it received for its fiscal 2004 report. If an agency receives a higher score, that score is a reliable indicator that the quality of its report has indeed improved.

Several factors helped ensure that the scoring criteria were applied more stringently this year. The same Mercatus Center research team evaluated the fiscal 2005 reports as evaluated the fiscal 2003 and 2004 reports. The team cross-checked the 2005 evaluations against those done in 2004 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 fiscal 2005 and fiscal 2004 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. In a number of cases, advisors thought the research team's initial scores were too lenient! Final scores thus reflect a careful review to ensure that the results of the scoring process are consistent with the goal of raising standards.