Introduction

This year, America celebrates the 400th anniversary of the settlement of Jamestown—a colony that almost perished due to a crisis of accountability. The colonists faced starvation because many refused to work. The "gentlemen" believed physical labor was beneath their dignity, and the less gentle preferred to pilfer tools, guns, and other items to barter with the natives for food. Neither group was held accountable for the negative consequences of its behavior.

In 1608, Captain John Smith became president of the colony's governing council. His solution to the accountability problem is legendary: "You must obey this now for a Law, that he that will not worke shall not eate (except by sickness he be disabled) for the labours of the thirtie or fortie honest and industrious men shall not be consumed to maintaine an hundred and fiftie idle loyterers."[1]

Smith pioneered performance reporting. "He made also a Table, as a publicke memorial of every mans deserts, to incourage the good, and with shame to spurre on the rest to amendment. By this many became industrious, yet more by punishment performed their businesse, for all were so tasked, that there was no excuse could prevaile to deceive him . . . "[2]

Smith employed these seemingly harsh accountability measures not just to reward those who did work, but also for a significant public purpose: to provide for those who could not care for themselves.

> You cannot deny but that by the hazard of my life many a time I have saved yours, when (might your owne wills have prevailed) you would have starved; and will doe still whether I will or noe; But I protest by that God that made me, since necessitie hath not power to force you to gather for your selves those fruits the earth doth yeeld, you shal not onely gather for your selves, but those that are sicke . . . The sick shall not starve, but equally share of all our labours; and he that gathereth not every day as much as I doe, the next day shall be set beyond the river, and banished from the Fort as a drone, till he amend his conditions or starve.[3]

Positive outcomes quickly ensued: "This order many murmured was very cruell, but it caused the most part so well to bestirre themselves, that of 200 (except they were drowned) there died not past seven."[4]

Though predating the Government Performance and Results Act by 385 years, Smith's initiative satisfies many of the prerequisites for successful performance management. Beneficial outcomes were well defined, performance measures were clear, causation was well understood, and performance was transparently linked to consequences. Performance reporting was public, accessible, jargon free, and verifiable.

[1] Philip H. Barbour, ed., *The Complete Works of Captain John Smith (1580-1631)*, vol II. (Chapel Hill: University of North Carolina Press, 1986): 208.

[2] *Ibid.*, 208.

[3] *Ibid.*, 213-14.

[4] *Ibid.*, 214.

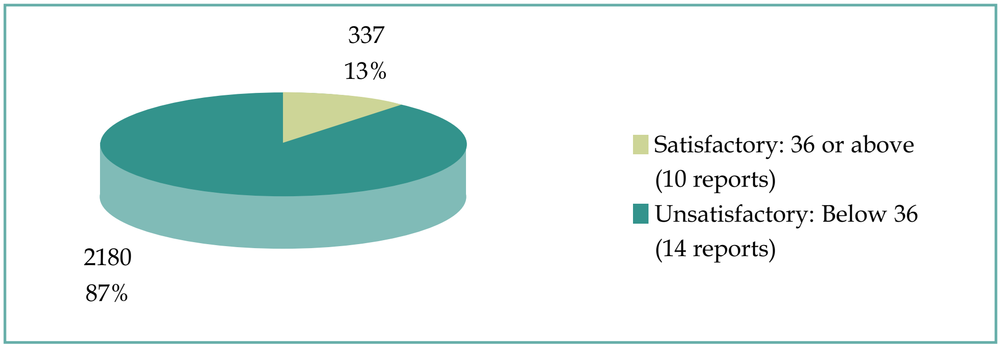
Today's federal government is rather more complex than the original Jamestown settlement. That makes effective accountability more difficult, but the same basic principles apply: define outcomes, define measures, understand how activities affect outcomes, and link outcomes to costs and consequences.

That's the genius of the Government Performance and Results Act (GPRA). Enacted in 1993, this legislation 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.[5]

Federal agencies are supposed to be accountable to elected policy makers. Elected policy makers are supposed to be accountable to citizens. Surely the GPRA-mandated disclosure of goals and results is a first crucial step toward both types of accountability.

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 issued by the 24 agencies covered by the Chief Financial Officers Act (CFO Act). These agencies accounted for 99 percent of federal outlays in fiscal 2006.[6] The 24 comprise all the Cabinet departments plus the largest independent agencies.

FIGURE 1: FISCAL 2006 SPENDING ($BILLIONS) COVERED BY SATISFACTORY DISCLOSURE



[5] 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?* (Arlington, VA: Mercatus Center at George Mason University, 2006), 21-22, http://www.mercatus.org/publications/pubid.2265/pub_detail.asp.

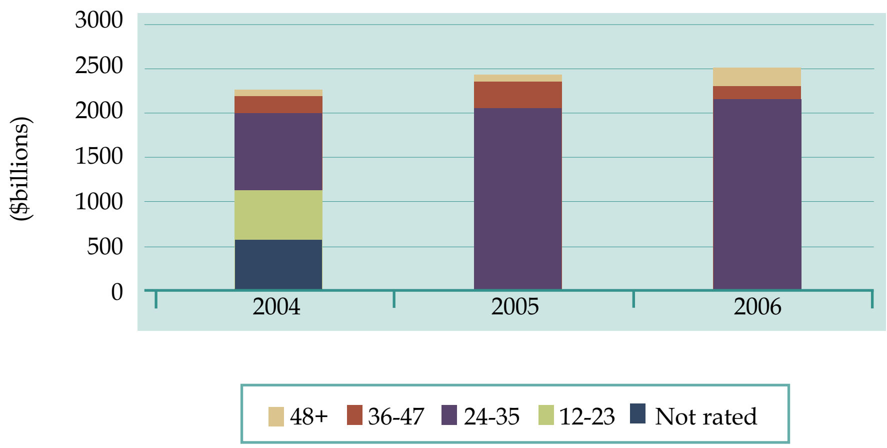
[6] The principal parts of government not included in these 24 agencies are the judiciary, the legislative branch, the executive office of the president, and the independent agencies not among the 24 CFO Act agencies. See *Budget of the United States Government, Fiscal Year 2008,* "Historical Tables," Table 4.1, 78. Outlays for these agencies actually exceed the "total outlays" figure, but they account for 99 percent of total outlays plus undistributed offsetting receipts.

The quality of disclosure in these reports has improved, but the improvement has not been as rapid or as widespread 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 2006, after eight years of reporting, only 13 percent of non-interest federal spending by the 24 agencies was covered by reports with total scores of "satisfactory" or better in our evaluation.[7] Though they covered only one-seventh of non-interest expenditures, these reports came from 10 of the 24 CFO Act agencies, as Figure 1 shows.

Figure 2 suggests that the best reports have improved their scores. More spending is now covered by reports with scores of very good (48 or better).[8] Nevertheless, the percentage of expenditures covered by satisfactory disclosure remains virtually unchanged between fiscal 2005 and 2006, and more than $2 trillion of the budget is still covered by unsatisfactory disclosure.

The four public benefits criteria most directly measure the quality of disclosure of outcomes. Figure 3 shows that in fiscal 2006 about 8 percent of the federal budget was covered by reports that earned a satisfactory score

Figure 2: Most spending covered by disclosure scoring below satisfactory (below 36)



[7] 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 page 9.

[8] 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 page 9.

on the public benefits criteria—12 or better. This is virtually the same as in fiscal 2005. However, almost all of the best reports actually scored 16 or better, thus earning a score of very good on the public benefits criteria. As Figure 4 shows, about $190 billion of non-interest spending, or 7.6 percent, was covered by reports rated very good on the public benefits criteria. In contrast, just 3 percent of spending was covered by reports earning a very good public benefits score in fiscal 2005.

Figure 3: Fiscal 2006 spending ($billions) covered by reports with satisfactory public benefits score (12 or above out of 20)

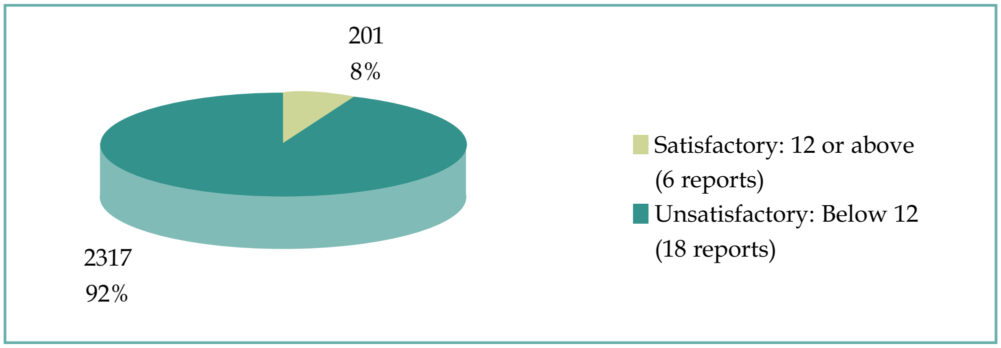
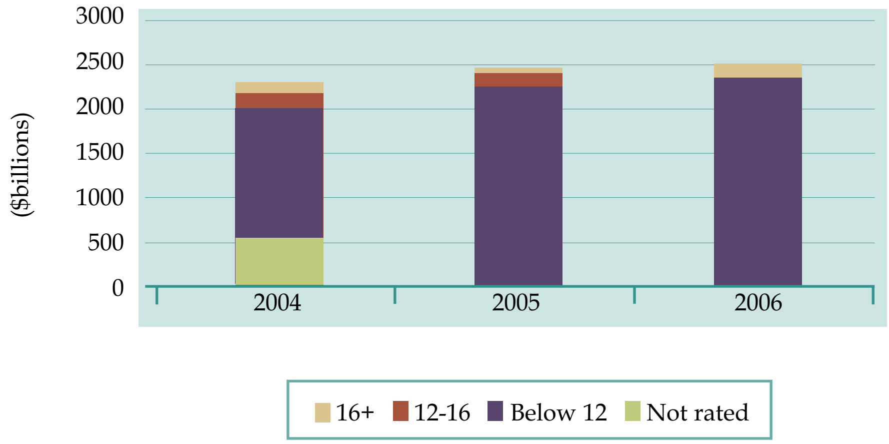


Figure 4: Most spending covered by public benefits disclosure scoring below satisfactory (below 12)



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 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.

Each agency's performance report necessarily addresses different audiences. Some readers are "stakeholders" who have expertise in the agency's work. They 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. From our perspective, though, the most important stakeholders are the ordinary citizens who pay the bills and deserve to know what the agency accomplished. A report will not score 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.

Reports that score high on our evaluation 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 tend to score well across the board. Reports that consistently score poorly do little to inform ordinary members of the public about important outcomes. Reports ranking in the middle may serve some audiences well, but they 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 lay person can follow and understand;
- present a set of performance metrics that captures important public outcomes that 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; and
- 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. 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 policy makers and the public understand what citizens are paying to achieve various outcomes.

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?

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. Thus reports must highlight achievement of results; otherwise, they will not inform the public of the success or failure of the 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 reports should inspire confidence in the agencies' abilities to enhance citizens' quality of life commensurate with the resources they have entrusted to them. Among the factors that give such confidence is tangible evidence that the agencies are using performance and financial data to improve management of their programs.

What Did the Agencies Know, and When Did They Know It?

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

In April 2006, when we released the fiscal year 2005 Scorecard, we also released a guidance document that explained how the best practices in the fiscal 2005 performance reports would affect the scoring of fiscal 2006 performance reports.[9] We have incorporated the guidance into the "Strongest and Weakest Scores" section of this Scorecard.

For fiscal 2006, 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 2006, 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 Internet by December 15 to earn credit on the first transparency criterion. A follow-up e-mail reminder was sent to the GPRA contacts on November 13. All but one of the agencies had their reports accessible on the Internet 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 iterated our goal of continuing to improve 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.

[9] "Annual Performance Report Scorecard Criteria Guidance" (Mercatus Center at George Mason University, April 2006), http://www.mercatus.org/publications/pubid.2264/pub_detail.asp.

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 for 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 criteria identifies 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 standard for best practice, but best practices should not be confused with perfection. We expect agency reporting practices to improve 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 (transparency, public benefit, and forward-looking leadership) separately, readers who believe that one factor is more

What do the scores mean?

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

4 Very Good
- Exceeds expectations
- Has potential to be 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 standard for adequate disclosure
- Shows no processs or plans to overcome problems
- Omits critical information

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, which deals 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 Health and Human Services, 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 produce 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 explain its results in understandable and meaningful terms 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 evaluate how the agencies 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. We 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

Our research team employed the same criteria to assess the fiscal 2006 agency reports that we used to evaluate prior year reports. In each succeeding year however we have tightened our evaluation standards. There are two reasons for this approach. One, the highest achievable quality is unlimited because creative innovators can always find ways to improve reporting practices and set new standards. Two, each year agencies have an opportunity to learn from 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 to 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 2006 report in order to receive the same numeric score it received for its fiscal 2005 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 help ensure that the scoring criteria are applied consistently from year to year. The same Mercatus Center research team has evaluated the reports for the past four years. The team cross-checked the 2006 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 fiscal 2006 and fiscal 2005 reports to ensure that differences in the actual contents of the reports justified any scoring differences across years. 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.