No. 09-03
March 2009

WORKING PAPER

HAS GPRA INCREASED THE AVAILABILITY AND USE OF PERFORMANCE INFORMATION?

By Jerry Ellig

MERCATUS CENTER
GEORGE MASON UNIVERSITY

"Has GPRA Increased the Availability and Use of Performance Information?"

Jerry Ellig[1]
Senior Research Fellow
Mercatus Center at George Mason University
jellig@gmu.edu

March 10, 2009

Abstract

The underlying logic of the Government Performance and Results Act (GPRA) suggests that programs should be evaluated based on empirical evidence that they actually produce the intended outcomes. This study applies the same logic to GPRA itself, investigating empirically whether GPRA may have increased the availability and use of performance information in federal agencies.

Better GPRA performance reporting is correlated with greater availability and use of several kinds of performance information by federal managers in the programs and operations they supervise. The results are statistically significant and relatively large. Correlations are especially significant for types of activities GPRA sought to encourage, such as output and outcome measures and use of performance information to allocate resources, set priorities, and develop measures and goals. These findings are consistent with the theory that GPRA has indeed prompted improvements in the availability and use of performance information in the federal government.

[1] The author would like to thank Thomas Beall at the Government Accountability Office for furnishing spreadsheets of GAO survey data and explaining the GAO findings in greater depth. I would also like to thank Mark Adams, Christina Forsberg, and Stefanie Haeffele-Balch for research assistance and Patrick McLaughlin and Richard Williams for helpful comments.

Introduction

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. The underlying logic of GPRA suggests that programs should be evaluated based on empirical evidence that they actually produce the intended outcomes. This study applies the same logic to GPRA itself, investigating empirically whether GPRA may have led to increased availability and use of performance information.

I examine the relationship between GPRA and performance measurement by combining two data sources: Government Accountability Office (GAO) surveys of federal managers (aggregated by agency) and the agency scores on the Mercatus Center's *Performance Report Scorecard*. Combining data from these two sources makes it possible to assess whether the quality of agencies' GPRA initiatives affects the availability and use of performance information reported by managers.

GPRA requires agencies to produce plans and reports, but they don't have to be good and they don't have to be used[2] (Brito and Ellig forthcoming). Previous scholarship offers many reasons GPRA might not spur the creation or the use of performance information to the extent its supporters hoped. Radin (2006) points out numerous barriers to effective implementation of GPRA. Tullock's (2005) and Downs's (1967) theories of bureaucratic behavior imply that implementation depends on whether agency officials believe performance management is important to their superiors—ultimately, Congress and the president. Agency theory, which undergirds public management reforms in Westminster countries like New Zealand and the United Kingdom, suggests that implementation of performance management will occur if policymakers align incentives of government agencies and individual officials so that they are rewarded for producing the desired outputs or outcomes and penalized when they fail to do so (Scott, Ball, and Dale 1997). Thus, many barriers could thwart the creation and use of performance information, even though GPRA is the law of the land.

Despite these potential barriers, a cursory glance suggests that GPRA may have had some positive effects on agency performance measurement. Periodic GAO surveys reveal that federal managers increasingly report that their agencies have performance measures, including output and outcome measures (Steinhardt 2008, 4). The Association of Government Accountants evaluates agencies' annual performance and accountability reports, awarding a Certificate of Excellence in Accountability Reporting to exceptional reports. The number of reports receiving this award has risen from two in 1998 (the first year of awards) to five in 2000 and 17 in 2007.[3] In addition, the Mercatus Center's annual *Performance Report Scorecard* shows that the average quality of agencies' annual performance reports improved by 17 percent between 1999 and 2006, in spite of the fact

[2] In Julnes and Holzer's (2001) terms, GPRA mandated at least the appearance of "adoption" but not "implementation."

[3] http://www.agacgfm.org/performance/cear/cearprioryear.aspx.

that the evaluators tighten the scoring criteria each year to reflect new best practices[4] (McTigue, Wray, and Ellig 2008, 17).

Yet even these data suggest some caution is in order. While GAO finds an increase in the availability of performance measures, it reports little improvement in the extent to which performance measures are actually used for significant managerial decisions, such as allocating resources, setting priorities, or setting job expectations (Steinhart 2008, 6). What's more, it is not clear whether even the increased availability of performance measures in the GAO surveys is a result of GPRA. The GAO surveys ask federal managers whether they have specific types of performance measures for their programs and operations, and whether they use performance information for various purposes in their programs and operations. When federal managers report that they have performance measures for their programs, such measures may be at a level far removed from, and maybe even unrelated to, the over-arching measures articulated in the agency's GPRA documents. Managers' positive responses may simply indicate that they have continued to develop and use measures suited to their programs, independent of whatever the agency is doing in response to GPRA.

GAO has identified numerous specific instances in which performance information has affected the decisions of managers in agencies (GAO 2005, Mihm and White 1999). Another way to explore the general relationship between GPRA and performance measurement is to correlate differences in the quality of an agency's GPRA initiatives with differences in the availability and use of performance information reported by federal managers. If managers in agencies with better GPRA initiatives are more likely to report that they have and use the types of performance measures envisioned by GPRA, then we can be more confident that GPRA is responsible for the observed improvement in the availability and use of performance information.

Table 1 summarizes the relevant aspects of GPRA, the GAO surveys, and the Mercatus Center's *Performance Report Scorecard*. The 24 agencies covered by the Chief Financial Officers Act, which are the subject of the GAO surveys and the Mercatus *Scorecard*, accounted for 96 percent of federal outlays in fiscal 2007 (McTigue, Wray, and Ellig 2008, 11).

[4] Scores fell somewhat for 2007, largely due to some agencies' difficulties implementing a new "pilot" format. The average score for 2007 was 11 percent higher than 1999.

Table 1: GPRA, the GAO surveys, and the Mercatus *Scorecard*

	Government Performance and Results Act (GPRA)	**Government Accountability Office surveys**	**Mercatus Center Performance Report Scorecard**
Coverage	Applies to virtually all federal agencies	24 federal agencies subject to the Chief Financial Officers' Act	24 federal agencies subject to the Chief Financial Officers' Act
Reporting Period	Requires agencies to produce multi-year strategic plans, annual performance plans, and annual performance reports	Conducted approximately every 3 years Only the 2000 and 2007 surveys were large enough to calculate valid responses for each agency	Conducted annually since fiscal 1999
Scope	Strategic plans must include performance measures Performance measures must include outputs and outcomes Performance plans must include performance goals Performance reports must report on goals and measures	Among other topics, federal managers were asked: (1) Whether they have specific types of performance measures, such as outcome, output, or quality measures (2) Whether they use GPRA strategic goals for various purposes, such as allocating resources or developing measures (3) Whether they use performance information for specific purposes, such as allocating resources or setting employee job expectations	Evaluation of the quality of agencies' annual GPRA performance reports Includes evaluation of: (1) Transparency: How easy is it to find the report, understand the report, and validate the data? (2) Public Benefits: How well does the report explain the outcomes the agency seeks to achieve and demonstrate how the agency has affected those outcomes? (3) Leadership: What evidence demonstrates that the agency's leadership uses performance information to manage the agency?

To the best of my knowledge, the only previous studies undertaking a general statistical analysis of GPRA's effects were two articles by Rainey and Han Chun in 2005. Rainey and Han Chun (2005a) developed several measures of "organizational goal ambiguity" and used these measures to evaluate 115 federal agencies. They measured "evaluative" ambiguity by assessing how outcome-oriented each agency's GPRA goals and measures are. Higher levels of evaluative ambiguity were associated with lower agency scores on managerial effectiveness, customer service orientation, productivity, and work quality in a 2000 survey of federal employees administered by the National Partnership for Reinventing Government (Rainey and Han Chun 2005b). In other words, federal employees perceive that their agencies are better managed, more customer-focused, more productive, and produce higher-quality work when GPRA goals and measures are more outcome-oriented. This study employs a similar method, correlating outside expert evaluations of the quality of agencies' GPRA products with surveys of people who work in the agencies.

1. Does theory suggest GPRA was likely to succeed?

Previous scholarship provides mixed predictions on whether GPRA could be expected to improve the availability or use of performance information. Reasons for skepticism can of course be found in the work of self-professed GPRA skeptics within the field of public administration. But some of the economic theories of organization that are often associated with public administration's renewed emphasis on performance also suggest that GPRA could be expected to succeed only under specific circumstances. The most pertinent are principal-agent theory and public choice.[5]

1.1. Skeptics

Numerous practical difficulties could make creation of reliable performance information envisioned by GPRA difficult or impossible. Outcome information may not be readily available, some data may be biased, multiple conflicting goals make agreement on measures difficult, cause-effect relationships are often unclear, baseline data may not exist, and key information may not be quantifiable (Radin 2006, 207–09).

Institutional factors also hamper both the creation and use of performance information. Unlike a parliamentary system, the design of the U.S. federal government divides power, and even within the executive and legislative branches different entities have their own perspectives and agendas. "As a result, it is very difficult—if not impossible—to craft a single government-wide effort that measures the performance of entities and also holds a single set of actors accountable for that performance" (Radin 2006: 148). Different kinds of programs also utilize different delivery strategies, such as direct federal activity, block grants, or research funding. Some kinds of programs are more amenable to output or outcome measurement than others (Roberts 2000; Radin 2006, 42–50). Thus progress in implementing GPRA should be uneven at best.

[5] Scott, Ball, and Dale (1997) explicitly identify agency theory and public choice economics as two critical influences on the wide-ranging management reforms adopted in New Zealand.

1.2 Principal-agent theory

It is also possible that GPRA would have limited effects because GPRA did not go far enough to align the incentives of agents (federal managers and employees) with principals (policy decisionmakers or voters). Public management reforms based on principal-agent theory, such as those undertaken in the Westminster countries, include measurement and transparent disclosure of outputs and outcomes, linkage of results with costs, a budget process that holds agencies responsible for results via purchase agreements, performance contracts that hold managers accountable for results, and devolution of control over inputs to the individuals held responsible for results. In New Zealand, government pay scales and lifetime job security were eliminated; pay and retention depend on performance. Scott, Ball, and Dale noted in 1997 (375):

> The New Zealand reforms tackled all parts of the public sector and all subsystems (personal, budgeting, reporting, and the like). The intent was to ensure that all the systems were strategically aligned and that they provided a consistent set of incentive signals. Indeed, it was only by tackling all parts and aspects of the public sector that significant gains were achieved. The same degree of comprehensiveness is not evident in the United States.

Kettl (1997, 454) notes that civil service reform has been an important part of performance management in other Westminster countries as well. The U.S. civil service system has had nowhere near the type of overhaul experienced in New Zealand. McNab and Melese (2003, 81–82) emphasize that effective performance budgeting requires changes in supporting budget processes and agency incentives that have not occurred in the United States. GPRA focuses on performance measurement, clarifying performance expectations, and performance budgeting. But it is doubtful that, in the majority of cases, the appropriations process holds agencies accountable either for outputs or outcomes (see, e.g., Curristine 2002, 42). Thus, the United States has not adopted the kind of comprehensive overhaul of incentives in public service that principal-agent theory suggests is necessary to maximize performance.

1.3 Public choice

Public choice analysis of bureaucracy begins with the assumption that government managers are neither selfless mandarins nor corrupt boodlers. Rather, they are similar to their reasonably intelligent fellow citizens in that the choices they make reflect the incentives and constraints they face (Niskanen 1994, 271). In a representative democracy, the extent of effective GPRA implementation should depend on the strength of managers' and legislators' incentives to care about performance.

Classic works by Tullock (2005) and Downs (1967) note that individuals in a bureaucracy achieve career advancement by performing their roles to the satisfaction of their superiors. It follows from this that agency managers can be expected to adopt and use performance measures if they believe that is what their superiors (the political appointees and, ultimately, Congress and the president) want them to do. Tullock (2005,

205–09) notes that measurement of results, where possible, improves the likelihood that bureaucracies will accomplish the tasks that policymakers intended them to accomplish. Even in cases where measurement of outcomes is difficult or impossible, his theory suggests that officials will do their best to implement it as well as they can, or at least create the appearance of implementation, if they believe that is what their superiors want. This theory predicts that top federal managers will develop and implement performance measurement and management if it is clear that Congress and the president want them to. Managers further down the ranks will take their cues from top management.

Congress and the president may have differing incentives to emphasize performance. Niskanen (1994) notes that members of Congress face conflicting incentives to monitor agency performance. On the one hand, the members of Congress serving on a given committee tend to be "high demanders" of the services provided by the agencies the committee oversees. They would like to see the services provided, and other things equal, they would like to see the services provided efficiently. However, monitoring of agencies is subject to a free rider problem because many of the benefits from monitoring—more, better, or lower-cost services—accrue to taxpayers and service beneficiaries in other districts. In contrast, members of Congress capture the full political benefits of performing services for constituents in their districts. Since their time is scarce, they face strong political incentives to under-provide monitoring of agency outputs, outcomes, and efficiency, in favor of performing constituent service.

The president is more likely than members of congressional committees to represent the "middle-demand" voters, since the president is elected by the entire nation and is more likely than individual committee members to be chosen by the nation's median voters (Niskanan 1994, 227). In that case, the president has stronger incentives than individual legislators to avoid supplying public services past the point where additional benefits equal the additional costs—which suggests the president may have stronger incentives to monitor agencies for efficiency than individual legislators.

The observed behavior of the president and Congress seem generally consistent with Niskanen's theory. The Clinton administration emphasized the importance of performance in its "reinventing government" initiative. During the G.W. Bush administration (the period covered by this study), the Office of Management and Budget (OMB) also signaled that performance management is important. OMB devoted substantial resources to performance measurement and management by developing extensive GPRA guidance for agencies and through the development of the Program Assessment Rating Tool (PART). PART attempts to apply performance management principles at the level of individual programs. The Bush administration tended to propose budget increases for programs that received good PART evaluations and decreases for programs judged ineffective or for which good results information was not available (Norcross and Adamson 2007, 25). In some cases, the administration cited PART assessments as its reason for proposing program terminations or funding reductions, but this was a fairly "noisy" signal, since budget cuts or increases also occurred for many other reasons (Gilmour and Lewis 2006; Norcross and Adamson 2007, 29–30). Nevertheless, OMB's desire to see improvements in performance measurement and management was clearly articulated. There was also some degree of linkage between demonstrated outcomes and budgetary consequences in the executive branch.

In general, Congress displayed less interest than the president. In the early years of GPRA, House Majority Leader Dick Armey took an active role in assessing agencies' strategic plans and performance plans. Apparently few appropriators or committee chairs shared his enthusiasm. In her study of GPRA implementation by the Department of Transportation, for example, Curristine (2002, 42) notes, "Indications from interviews with appropriators show that they will not use performance measures in making funding decisions on highways."

Similarly, there is little evidence that Congress used PART as a budgeting tool. In fiscal 2007, Congress did tend to give funding increases to programs with good PART ratings and decreases to ineffective or "results not demonstrated" programs, though not to the same extent as the president proposed (Norcross and Adamson 2007, 28). An analysis of committee reports in the 109th Congress revealed that only about 6 percent of them had PART-related content, which leads the authors to conclude that Congress used PART "on a limited basis." One subcommittee even banned departments under its jurisdiction from including PART information in its fiscal 2008 budget submission (Frisco and Stalebrink 10, 16).

1.4 Summation

Scholarship provides several reasons to believe that GPRA would go the way of Planning, Programming and Budgeting, Zero-Based Budgeting, and other failed management and budget reforms. Theory suggests that full implementation of GPRA as a management tool may be so difficult as to be impossible in many contexts, or perhaps the United States failed to adopt sufficiently sweeping changes in public managers' incentives and constraints to motivate widespread implementation.

GPRA might have a large effect if some blend of three motivations is sufficient to induce managers to implement and adopt performance management: (1) the mere fact that GPRA is the law, (2) guidance and pressure from the Office of Management and Budget, reflecting the wishes of the executive, and/or (3) agency officials' belief that it is "the right thing to do," which they would pursue regardless of outside pressure or incentives.

2. The Data

This study combines data from two sources: the Government Accountability Office's periodic surveys of federal managers on the existence and use of performance information, and the Mercatus Center's *Performance Report Scorecard* evaluations of agencies' annual performance reports.[6] Both data sets cover the 24 agencies subject to

[6] As an alternative measure of the quality of an agency's GPRA initiatives, I checked to see whether an agency received the Association of Government Accountants' Certificate of Excellence in Accountability Reporting (CEAR) in 2000 or 2007 and also counted the number of times each agency received this honor for 2000-2007 inclusive. In regressions similar to those reported below, this variable often had a positive coefficient but was not statistically significant at conventional levels (i.e., 90 percent or above). This probably occurred because agencies must volunteer for AGA's CEAR program, and so it is possible that some excellent reports did not receive recognition because they did not volunteer to be evaluated. It is comforting to find that the CEAR variable had the same sign as the *Scorecard* variables.

the Chief Financial Officers' Act.[7] For 2000 and 2007, GAO surveyed a large enough sample of managers to permit calculation of valid response averages for each agency. Appendix I presents summary statistics on all data used to produce results reported in this paper.

2.1 The GAO data

Three groups of GAO survey questions are of interest:

(1) Six questions asking about the existence of performance measures (quality, customer satisfaction, efficiency, output, outcome, or any at all).

(2) Nine questions asking whether managers use performance information for various purposes in the programs or other activities for which they are responsible (allocating resources, setting priorities, adopting new approaches or changing work processes, coordinating with external organizations, refining performance measures, setting performance goals, setting job expectations, rewarding employees, managing contracts).

(3) Several "environmental" questions that can be used as control variables in regressions, such as one that asks managers to gauge the strength of agency leadership's commitment to performance management and one that asks managers whether they have sufficient information to judge the validity of the performance information they have.

Following the practice in numerous GAO reports, this study uses the percent of managers who responded "to a great extent" or "to a very great extent" as the measure of the strength of managers' affirmative responses to the questions.

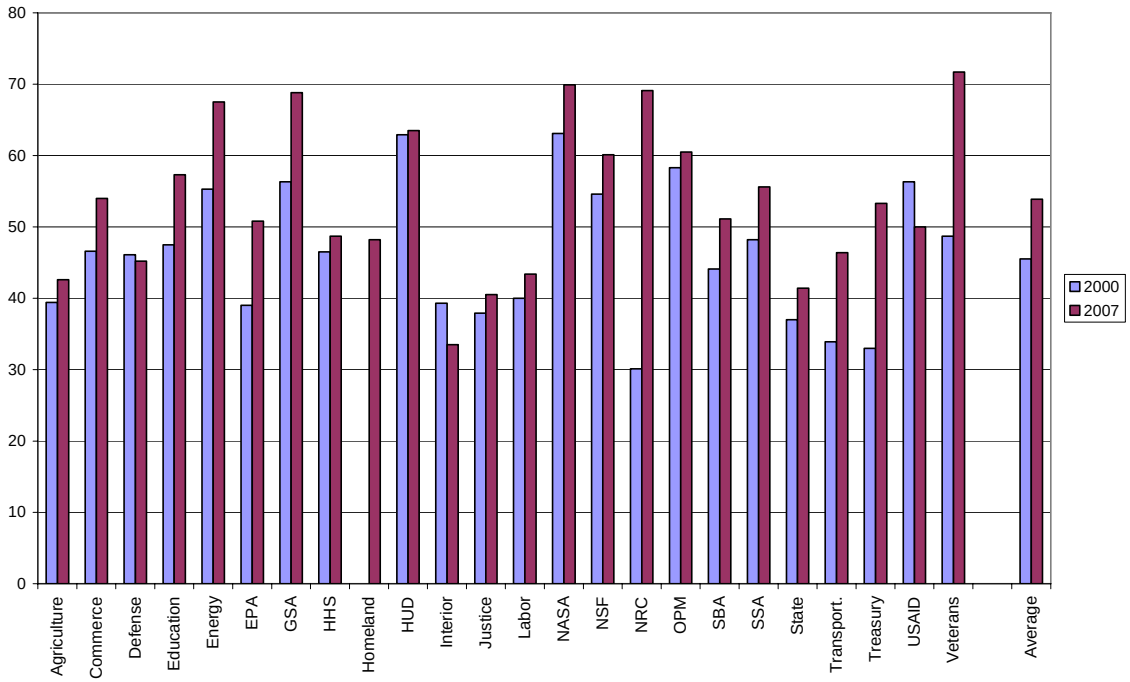
Figure 1 shows the percent of managers in each agency who report that they have outcome measures in 2000 and 2007 to a great or very great extent. It provides a good indication of the nature of the GAO survey data. The percentage of affirmative responses ranges from 28.1 (FEMA) to 63.1 (NASA) in 2000, and from 40.5 (Justice) to 71.1 (Veterans Affairs) in 2007. The average percentage of affirmative responses increased from 45.5 in 2000 to 53.9 in 2007.

GAO survey data measure managers' stated perceptions about the availability and use of performance information. The surveys, therefore, might not measure the actual extent to which performance information is available or used—either because some managers' perceptions could be inaccurate or because some managers might be prone to giving the answers they feel they are expected to give instead of stating their true perceptions. Nevertheless, the GAO surveys are the best measure I have seen that allows one to compare the availability and use of performance information across agencies and across

[7] In some cases, GAO broke out survey results separately for certain components of cabinet departments, such as the Federal Aviation Administration vs. the rest of the Department of Transportation. In these cases I computed a weighted average for the entire department, using the number of managers surveyed in sub-units as weights. Thomas Beall at the Government Accountability Office graciously furnished the totals for each sub-unit.

time periods, and there is no intuitively obvious reason that possible inaccuracies would make the data unsuitable for the purposes of this study.

Figure 1: Percent of managers saying their programs have outcome measures



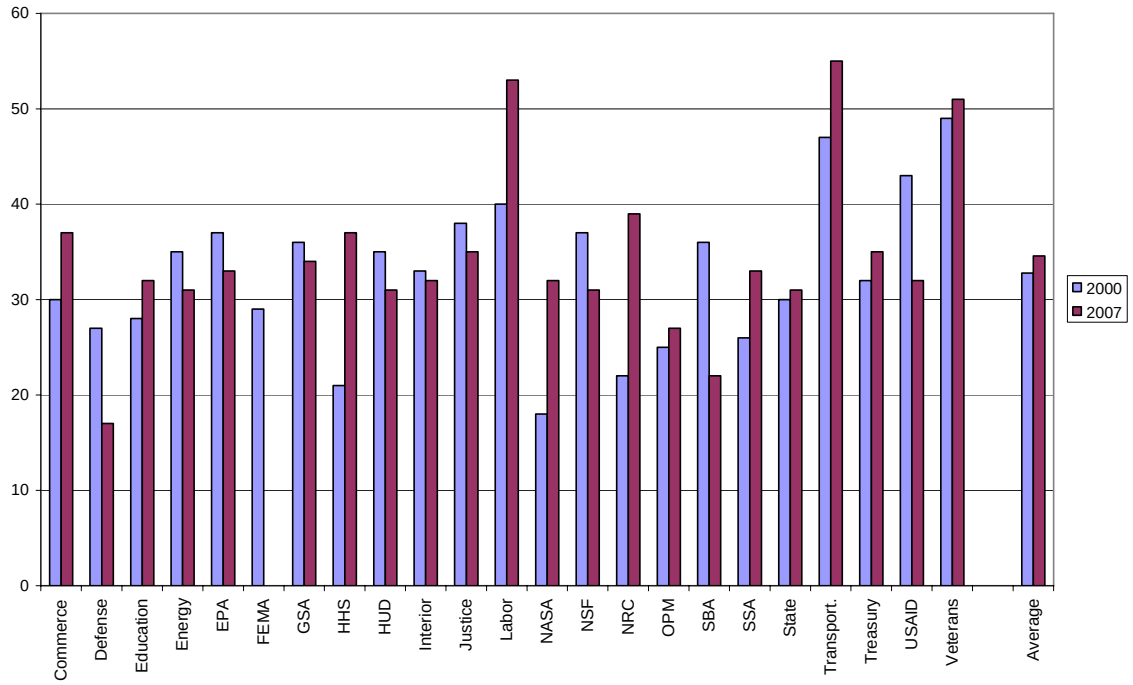
2.2 Mercatus *Scorecard* data

The Mercatus Center at George Mason University annually evaluates the quality of the annual performance reports produced by the 24 CFO Act agencies. A research team scores the reports on 12 criteria grouped into three categories: Transparency (Was the report easy for a non-specialist to find and understand?), Public Benefits (Does the report have outcome-oriented goals and measures that are linked to costs?), and Leadership (Is there evidence in the report that the agency's management uses performance information to guide decisions?) Scores on each criterion range from 1 (no useful content) to 5 (potential best practice), and so the total score on the report can range from 12 to 60 points.

Each year, the research team tightens the scoring criteria to reflect the previous year's best practices. As a result, improvement in an agency's score over time under-estimates the actual improvement in its report. An increased score definitely signifies improvement, but a static score implies that the agency's report has improved somewhat. A reduction in the score over time may mean that the quality of an agency's report has fallen, or it may just mean that the agency's report has not improved at as rapid a rate as other agencies.

Figure 2 shows the scores earned by each agency's report in 2000 and 2007.[8] They ranged from 18 out of a possible 60 (NASA) to 49 out of a possible 60 (Veterans Affairs) in 2000, and from 17 (Defense) to 55 (Transportation) in 2007. The average increased from 32.8 in 2000 to 34.6 in 2007. This study uses these scores as an indicator of the quality of an agency's GPRA initiatives. (Score data are from McTigue, Ellig and Richarson 2001 and McTigue, Wray, and Ellig 2008.)

Figure 2: Mercatus Scorecard scores



[8] The Department of Agriculture's performance report was not produced in time to be included in the Mercatus Center evaluation in 2000. The Federal Emergency Management Agency's report was evaluated in 2000 but not in 2007, because by then FEMA was part of the Department of Homeland Security. The Department of Homeland Security did not produce a report in 2000 because it did not exist then.

Do the 2 data sets measure the same thing?

One potential criticism of this paper might be that a link between Mercatus *Scorecard* scores and the GAO surveys should be expected because they measure the same thing. The Mercatus Scorecard, after all, evaluates the quality of the agency's GPRA measures and evidence in the performance report that the agency uses performance information. The GAO asks managers if they have and use performance information.

There is, however, a key difference between what the Mercatus *Scorecard* measures and what the GAO survey measures. The Mercatus *Scorecard* evaluates (among other factors) the quality of the agency-wide strategic goals and measures that appear in the agency's GPRA report. The Mercatus *Scorecard's* leadership evaluation focuses on top management's responses to major management challenges and significant initiatives to improve performance in the future.

The GAO survey asks federal managers whether they have specific types of performance measures *for their programs and operations*, and whether they use performance information for various purposes *in their programs and operations*.[1] There is, of course, no guarantee that these will measure the same thing the Mercatus *Scorecard* measures. The mere fact that performance information appears in the agency's GPRA report does not automatically mean that managers will pay attention to it or use it. In addition, the GAO survey will often implicate performance information at a much lower level than the agency's over-arching GPRA goals and measures. The survey asks managers whether they personally use performance information, which usually involves decisions taken at lower management levels than those examined in the Mercatus *Scorecard*. For these reasons, any correlations between the Mercatus *Scorecard* scores and GAO survey results likely indicate that GPRA has moved beyond a compliance exercise to drive development and use of performance information at multiple levels of the managerial hierarchy.

3. Trends in availability and use of performance data

Taking individual agencies as the unit of analysis, performance measures of all types were more available in 2007 than in 2000, as table 2 shows. Most agencies were measuring some kind of performance in both years; the percentage of managers who said they had some kind of performance measures for the programs or operations they were involved with increased from 87.3 in 2000 to 89.9 in 2007. Much larger percentages of managers reported improvements in the existence of particular types of measures for their programs or operations. The average proportion of managers reporting that their programs had specific types of performance measures to a great or very great extent increased by between 6 and 9 percentage points, depending on the measure. The largest and most statistically significant increases (at the 99 percent level) were for outcome and efficiency measures. Interestingly, the largest increase occurred in the category where GPRA provided unique new emphasis: outcomes.

Table 2 also reveals that in 2007, larger percentages of managers reported that they used the agency's GPRA strategic goals to a great or very great extent to set program priorities, allocate resources, adopt new programs or change program processes, and develop performance measures. The average proportion of managers reporting that they use the strategic goals to a great or very great extent for these purposes in their programs rose by between 7 and 9.4 percentage points, depending on the purpose. All differences are statistically significant at the 99 percent level.

Table 2: Percentage of managers answering "great extent" or "very great extent"

	2007	2000	Difference	T-statistic
Existence of performance measures				
Any kind (% answering yes)	89.9	87.3	2.6	1.73*
Output	59.4	53.4	6.0	1.92*
Efficiency	45.2	36.2	9.0	3.34***
Customer satisfaction	43.2	36.6	6.6	1.97*
Quality	43.1	36.9	6.2	1.98*
Outcome	53.9	45.5	8.4	2.80***
Uses of GPRA strategic goals				
Setting program priorities	79.5	72.5	7.0	3.89***
Allocating resources	73.3	65.2	8.1	4.67***
Adopting new programs/changing processes	72.5	64.9	7.6	4.03***
Developing or refining performance measures	67.2	57.8	9.4	4.00***
Uses of performance information in their programs				
Setting priorities	53.2	46.4	6.7	2.48***
Allocating resources	50.5	45.6	4.9	1.99*
Adopting new approaches/work processes	51.3	42.5	8.8	3.18***
Coordinating with external organizations	45.4	35.7	9.7	4.17***
Refining program performance measures	44.5	38.2	6.3	2.37***
Setting or revising performance goals	50.2	43.3	6.9	2.67***
Setting job expectations for employees I manage	55.0	42.8	12.3	4.80***
Rewarding employees I manage or supervise	52.7	43.6	9.1	3.19***
Managing contracts	28.8	24.2	4.5	1.51

Statistical significance: *90% **95% ***99%

Use of performance information also increased. Table 2 shows that higher percentages of managers reported that they used performance information in their programs or operations to set priorities, allocate resources, adopt new programs or change program processes, coordinate with external organizations, refine performance measures, set or revise program goals, set or revise employees' job expectations, and reward employees. The average proportion of managers reporting that they use performance information to a great or very great extent for these purposes rose by between 4.9 and 12.3 percentage points. All of these differences are statistically significant, usually at the 99 percent level. Performance information was not used more extensively to manage contracts in 2007 than it was in 2000.

These figures are consistent with GAO's most recent finding of a statistically significant change from 1997 to 2007 in the percentage of managers reporting that they have various kinds of performance measures. However, GAO did not usually find similar improvements in the percentage of managers who report that they use performance measures for specific purposes. The only use of performance information that showed

statistically significant improvement from 1997–2007 was to reward employees (Steinhardt 2008, 4–7).

There are two possible reasons the results on use of performance information differ in this paper and the most recent GAO survey. First, GAO compared 1997–2007, and for some unknown reason the percentages of positive responses in 1997 were usually higher than in 2000 (Private communication). Second, this paper weights the managers' responses differently than GAO's report. GAO usually reports averages for the entire sample of federal managers, so each manager counts as one observation. In contrast, the figures reported above use each agency as one observation. The average for each agency counts equally. Thus, this study gives greater weight to the responses from managers in smaller agencies. It is doubtful that this accounts for the difference, however, as the regression analyses performed for this paper found no statistically significant relationships between changes in the GAO survey results and agency size as measured by outlays or change in outlays.[9]

4. GPRA's role in the improvement

The fact that agencies were more likely to have and use performance information in 2007 than in 2000 does not necessarily mean that GPRA deserves credit for the improvement. As section 1 indicated, there are solid theoretical reasons for expecting that GPRA would not have much of an effect. Rather than reflecting the unique influence of GPRA, the observed improvement may have been a legacy of the Clinton administration's "reinventing government" initiative, a response to the G.W. Bush administration's management initiatives that were independent of GPRA, a learning curve effect, or a trend that occurred for some other exogenous reason.

We could be more confident that GPRA caused some of the improvements indicated by the GAO survey results if the quality of agencies' GPRA initiatives is positively correlated with the percentages of managers who have and/or use performance information. That relationship is tested in this section using two types of ordinary least squares econometric specifications: a pooled 2000/2007 data set with control variables, and a "difference" approach that regresses the 2007–2000 difference in GAO survey responses on the 2007–2000 difference in *Scorecard* scores.

4.1 Pooled data with control variables

The first specifications regressed the dependent variables of interest—the GAO survey responses for 2000 and 2007—on the agency's *Scorecard* score in those years.[10]

[9] In a series of regressions not reported in this paper, the only results even close to a statistically significant relationship occurred between outlays and the change in the percent of managers reporting that they have customer satisfaction measures or use performance information to coordinate with external organizations or set program goals. This relationship was only significant at the 90 percent level, and significance fell below this level when a control variable measuring the change in top leadership's commitment to performance management was added to the regressions.

[10] This approach yields 46 observations. The Federal Emergency Management Agency could not be used in 2000 because the PART variables controlling for various types of programs as a percent of its budget were

Potential control variables included a dummy variable for the year and a set of variables that account for agency-specific factors that might influence the percent of managers offering positive responses. These latter variables included commitment of the agency's leadership to performance management, the validity of performance information as perceived by managers (both measured by the responses to GAO survey questions on those topics), agency size (measured by outlays), percent of the agency's spending that is discretionary rather than mandatory, and agency ideology (measured by the Clinton-Lewis [2008] expert evaluation).

Control variables also included the percentage of the agency's budget devoted to regulation, research and development, competitive grants, and block and formula grants.[11] Several authors (Radin 2006, Roberts 2000, Han Chun and Rainey 2005a) suggest that these types of programs are different from direct federal programs that provide services; agencies may have a more difficult time articulating or measuring results for these programs than for others. Consequently, managers in agencies where these types of programs are larger might be expected to have or use performance information less extensively.

Some of the control variables were rarely statistically significant and hence were dropped from the equations. These include perceived validity of the performance information, outlays, percent of discretionary spending, and agency ideology.

4.1.1 Availability of performance information

The *Scorecard* variable has some degree of statistically significant correlation with the availability of three types of performance information: outcome measures, output measures, and efficiency measures.[12] Table 3 shows two alternative specifications for each type of measure. The effect of the year dummy variable is not clear. Year 2007 indicates that the percentage of positive responses was lower in 2007, after controlling for other factors. The interaction variable Score*year, however, suggests that Score might have a smaller effect in 2007. The equations using Score*year provide a slightly better fit, and in any case, the results in both specifications are quite similar.

---

not available. The Department of Agriculture could not be used in 2000 because its report was not produced in time to be included in the Mercatus Center evaluation in 2000.

[11] These percentages are calculated for fiscal 2008 from a spreadsheet produced by the Office of Management and Budget as part of its Program Assessment Rating Tool, available via www.expectmore.gov. Because PART gradually expanded its scope to review most federal programs, the spreadsheet produced in fiscal 2008 contains the most comprehensive list of programs. Since only 2008 and 2009 spending data are available for the complete list of programs, the percentages spent on various types of programs only approximately control for differences across agencies in 2000 and 2007.

[12] In the interest of conserving space, regressions for customer, quality, or "any" type of performance measures are not reported since Score was not statistically significant.

Table 3: GPRA report quality correlated with availability of performance measures

46 observations	**Outcome**	**Outcome**	**Output**	**Output**	**Efficiency**	**Efficiency**
Score	.36 [1.91*]	.30 [1.84*]	.47 [2.63**]	.31 [1.94*]	.32 [1.76*]	.29 [1.85*]
Score*year	-.10 [-.99]		-.25 [-2.59**]		-.04 [-.37]	
Year 2007		-3.4 [-.99]		-7.16 [-2.16**]		-.67 [.20]
Leadership	.70 [4.25***]	.69 [4.40***]	.82 [5.31***]	.76 [5.01***]	.61 [3.81***]	.58 [3.84***]
Comp. grant	.13 [1.93*]	.13 [1.96*]	.09 [1.34]	.09 [1.31]	-.06 [-.91]	-.06 [-.92]
Block grant	-.06 [-.94]	-.06 [-.99]	.01 [.21]	.003 [.05]	-.09 [-1.63]	-.10 [-1.69*]
Regulatory	-.12 [-1.89*]	-.12 [-1.93*]	.09 [1.45]	.08 [1.35]	-.09 [-1.40]	-.08 [-1.37]
R&D	.02 [.31]	.02 [.34]	-.17 [-2.89***]	-.16 [-2.68***]	-.10 [-1.69*]	-.10 [1.64]
Constant	-3.63 [-.27]	-1.14 [-.10]	-.06 [-.49]	2.59 [.23]	-3.92 [-.30]	-1.88 [-.17]
Adj. R-squared	.43	.43	.46	.43	.42	.42

Statistical significance: *90% **95% ***99%

Other noteworthy results include the very high statistical significance of leadership—a factor noted in GAO's own discussion of its survey results (Steinhardt 2008, 9–11). In some cases, the percentage of an agency's budget devoted to competitive grants, block grants, regulation, or research and development is also correlated, positively or negatively, with the availability of some types of performance information:

- Agencies with a higher percentage of their budgets devoted to competitive grants seem more, not less, likely to have outcome measures.

- Agencies with a higher percentage of block grants appear less likely to have efficiency measures.

- Agencies with a higher percentage of their budgets devoted to regulation have lower percentages of managers that say they have outcome measures.

- Agencies with more of a research and development focus have lower percentages of managers reporting they have output and, perhaps, efficiency measures.

4.1.2 Use of performance information

Two ways GPRA might affect use of performance information

Better GPRA initiatives →
Greater use of performance information
(Table 4)

Better GPRA initiatives →
More performance information available →
Greater use of performance information
(Table 5)

There are two ways to test whether the quality of an agency's GPRA initiatives leads to increased use of performance information. The first way takes the same approach as table 3, regressing the percentage of managers saying they use performance information on the *Scorecard* variable and control variables. Table 4 presents these results.

The second way starts with the hypothesis that agencies are more likely to use performance information if they *have* performance information. The results in table 3 show that better performance on the Mercatus *Scorecard* is associated with increases in the percentage of managers who say they have output, outcome, and efficiency measures. If greater availability of these measures is correlated with greater use of performance information, then GPRA may increase the use of performance information simply by making more performance information available. This hypothesis can be tested by regressing the percentage of managers saying they use performance information on the percentage of managers saying they have outcome, output, and/or efficiency measures.

When this was tried using all three measures, only outcome and output measures had a statistically significant correlation with the various uses of performance information. Hence, table 5 presents econometric results using only the outcome and output variables. All of the other control variables were included, to see whether any of these variables is correlated with uses of performance information after controlling for the availability of outcome and output measures.

The score of agency GPRA reports is positively correlated with five different uses of performance information: allocating resources, setting priorities, coordinating with external agencies, establishing measures, and setting goals.[13] Correlations are statistically significant at the 90 percent level and sometimes at the 95 percent level.

[13] As in table 3, results using the Year dummy variable are similar to the results using Year*score, but the latter provide a slightly better fit and the former are omitted to conserve space. One exception is the regression for coordination with external parties, where excluding any year variable leads to the best fit.

Table 4: GPRA report quality correlated with uses of performance measures

46 observations	**Allocate**	**Priorities**	**Coordinate**	**Coordinate**	**Measures**	**Goals**
Score	.35 [2.21**]	.40 [2.17**]	.27 [1.69*]	.28 [2.12**]	.34 [1.88*]	.34 [1.89*]
Score*year	-.14 [-1.66]	-.15 [-1.45]	.01 [.11]		-.11 [-1.12]	-.13 [-1.33]
Leadership	.57 [4.16***]	.70 [4.38***]	.53 [3.76***]	.54 [5.82***]	.60 [3.74***]	.68 [4.36***]
Comp. grant	.09 [1.63]	.03 [.45]	.13 [2.18**]	.13 [2.26**]	.12 [1.78*]	.12 [1.88*]
Block grant	-.05 [-.92]	-.001 [-.02]	-.07 [-1.48]	-.07 [-1.50]	-.03 [-.60]	.006 [.11]
Regulatory	-.10 [-1.95*]	-.04 [-.61]	-.03 [-.53]	-.03 [-.57]	-.05 [-.85]	-.04 [-.66]
R&D	-.02 [-.46]	-.21 [-3.48***]	-.11 [-1.99*]	-.01 [-2.14**]	-.11 [-1.72*]	-.06 [-1.09]
Constant	4.08 [.37]	-3.37 [-.26]	-1.03 [-.09]	-1.91 [-.24]	-9.07 [-.68]	-5.37 [-.42]
Adj. R-squared	.34	.36	.44	.45	.29	.34

Statistical significance: *90% **95% ***99%

As in table 3, leadership has a highly significant correlation with these uses of performance measures. Agencies with a higher percentage of their budgets devoted to competitive grants seem more likely to use performance information to coordinate with external parties, develop measures, and set goals. Agencies that do more regulation appear less likely to use performance information to allocate resources. Finally, the percent of budget spent on research and development has a negative and statistically significant correlation with use of performance information to set priorities, coordinate with external parties, and develop measures.

Table 5 shows how the quality of GPRA reports could be indirectly correlated with all nine uses of performance information in the GAO survey via GPRA's correlation with the availability of outcome and output measures. In almost all cases, the existence of outcome measures has a positive and statistically significant correlation with use of performance information. In three cases—setting priorities, changing programs or work processes, and coordinating with external parties—the existence of output measures is also correlated with use of performance information. Finally, use of performance information to set job expectations is positively and significantly correlated with output but not outcome measures. Since Score is positively correlated with outcome and output measures, the quality of an agency's GPRA initiatives is indirectly correlated with all uses of performance information.

Table 5: Existence of measures and uses of performance measures

46 observations	**Allocate**	**Allocate**	**Priorities**	**Priorities**	**Change**	**Change**
Outcome	.69 [8.81***]	.70 [7.92***]	.57 [4.34***]	.34 [2.88***]	.50 [3.64***]	.36 [2.46**]
Output		-.02 [-.18]		.56 [4.66***]		.35 [2.34**]
Score	.10 [1.07]	.11 [1.06]	.19 [1.22]	.01 [.08]	.09 [.55]	-.02 [-.13]
Score*year	-.08 [-1.46]	-.08 [-1.40]	-.09 [-1.06]	.03 [.44]	-.04 [-.45]	.04 [.39]
Leadership	.09 [.95]	.10 [.91]	.31 [1.90*]	-.003 [-.02]	.35 [2.08**]	.16 [.88]
Comp.grant	.002 [.06]	.003 [.07]	-.04 [-.78]	-.06 [-1.37]	.03 [.42]	.01 [.24]
Block grant	-.07 [-.24]	-.006 [-.22]	.03 [.63]	-.01 [.27]	-.03 [-.67]	-.05 [-.97]
Regulatory		-.02 [-.54]	.03 [.56]	-.05 [-1.03]	-.06 [-1.17]	-.11 [-2.01*]
R&D	-.04 [-1.25]	-.04 [-1.17]	-.22 [-4.44***]	-.12 [-2.64***]	-.15 [-2.92**]	-.09 [-1.60]
Constant	6.59 [1.03]	6.52 [1.00]	-1.30 [-.12]	-1.20 [-.19]	1.76 [.17]	1.76 [.17]
Adj. R-squared	.78	.78	.57	.72	.60	.60
46 observations	**Coordinate**	**Coordinate**	**Measures**	**Measures**	**Goals**	**Goals**
Outcome	.44 [3.56***]	.33 [2.52**]	.70 [5.50***]	.62 [4.56***]	.64 [5.43***]	.59 [4.52***]
Output		.25 [1.82*]		.12 [.88]		.11 [.82]
Score	.12 [.79]	.04 [.24]	.09 [.63]	.05 [.34]	.11 [.78]	.07 [.5]
Score*year	.05 [.68]	.11 [1.31]	-.04 [-.55]	-.02 [-.20]	-.07 [-.89]	-.04 [-.53]
Leadership	.23 [1.52]	.09 [.56]	.12 [.78]	.05 [.29]	.24 [1.66]	.18 [1.09]
Comp.grant	.07 [1.32]	.06 [1.19]	.08 [1.49]	.08 [1.40]	.04 [.75]	.03 [.67]
Block grant	-.05 [-1.13]	-.06 [-1.36]	.01 [.25]	.007 [.15]	.04 [.97]	.04 [.67]
Regulatory	.02 [.46]	-.01 [-.22]	.05 [1.05]	.03 [.66]	.04 [.76]	.02 [.41]
R&D	-.11 [-2.45]	-.07 [-1.34]	-.09 [-1.86*]	-.06 [-1.21]	-.08 [-1.71*]	-.06 [-1.1]
Constant	.56 [.06]	1.74 [.18]	-2.03 [-.20]	-1.45 [-.14]	-3.05 [-.32]	-2.53 [-.26]
Adj. R-squared	.57	.60	.59	.58	.62	.62

Table 5 cont.						
45 observations	**Job Exps.**	**Job Exps.**	**Reward**	**Reward**	**Contracts**	**Contracts**
Outcome	.35 [2.46**]	.17 [1.18]	.46 [2.86***]	.37 [2.09**]	.61 [3.65***]	.62 [3.32***]
Output		.45 [2.88***]		.22 [1.17]		-.03 [-.17]
Score	.06 [.35]	-.08 [-.48]	.08 [.43]	.01 [.07]	-.36 [-1.80*]	-.35 [-1.65]
Score*year	.08 [.87]	.18 [1.92*]	.03 [.27]	.07 [.67]	-.07 [.64]	.06 [.53]
Leadership	.39 [2.23**]	.15 [.85]	.25 [1.28]	.13 [.61]	-.10 [-.48]	-.08 [-.34]
Comp. grant	-.02 [-.32]	-.04 [-.61]	-.09 [-1.35]	-.1 [-1.45]	.16 [2.20**]	.16 [2.18**]
Block grant	-.03 [-.61]	-.05 [-.98]	-.07 [-1.19]	-.08 [-1.31]	.01 [.24]	.02 [.25]
Regulatory	-.04 [.66]	-.10 [-1.72*]	-.05 [-.76]	-.08 [-1.13]	.07 [1.02]	.07 [1.00]
R&D	-.2 [-3.70***]	-.12 [-2.18**]	-.18 [-3.03***]	-.15 [-2.12**]	.07 [1.05]	.06 [.83]
Constant	5.76 [.49]	7.82 [.72]	10.5 [.81]	11.5 [.89]	10.5 [.78]	10.39 [.75]
Adj. R-squared	.56	.63	.44	.45	.40	.39

Interestingly, the Score variable and other control variables rarely have a positive correlation with uses of performance information after controlling for the existence of outcome and output measures. The most consistent statistically significant correlation involves research and development, which has a negative and significant correlation with use of performance information to set priorities, change programs or processes, set job expectations, reward employees, and perhaps establish measures and goals.

4.2 Differences

In some regressions in tables 3–5, Score is correlated with availability and use of performance information only at the 90 percent level of statistical significance. This might be a sign of a weak correlation, or it may simply result from other agency-specific factors that the regressions fail to control for. A more comprehensive method of controlling for agency-specific differences is to regress the 2007–2000 *difference* in survey responses for each agency on the 2007–2000 *difference* in the agency's Scorecard score, plus possibly some control variables. Taking the differences controls for any unobserved agency-specific factors that were present in both years. Using the differences focuses the analysis on whether *improvements* in an agency's GPRA initiatives led to *improvements* in the adoption and use of performance information.

The main disadvantage of using differences is that it cuts the sample size down to 22 agencies. But if statistical significance of the correlations is high and results are similar to those in the pooled approach, we can be reasonably confident that the results are not merely an artifact of the small sample size.

4.2.1 Availability of performance information

Table 6 shows how improvement in an agency's GPRA initiatives is correlated with the change in the percentage of managers reporting that they have specified types of performance measures to a great or very great extent for their programs or operations. "Difference Scorecard Score" is the 2007 score minus the 2000 score; therefore, a positive number indicates improvement in the score. If GPRA report quality is positively correlated with availability or use of performance measures, the coefficient on the score variable should be positive. This is precisely what table 6 shows.[14]

Since Leadership emerged as such a strong factor in the pooled regressions, regressions were also run that included a control variable measuring the 2000–07 change in the percent of managers saying the agency's leadership is committed to performance management to a great or very great extent. This variable is sometimes significant, and it does not materially reduce the statistical significance of the Score variable.

Table 6: Difference in GPRA report quality and difference in availability of performance information

22 observations	**Outcome**	**Outcome**	**Output**	**Output**	**Customer**	**Customer**	**Any**	**Any**
Score	.47 [2.00*]	.52 [2.33***]	.40 [2.70***]	.42 [3.02***]	.43 [1.84*]	.45 [1.90*]	.28 [1.77*]	.24 [1.69]
Leadership		.35 [1.92*]		.20 [1.74*]		.15 [.78]		-.02 [-.15]
Constant	7.33 [3.62***]	1.71 [.49]	4.67 [3.71***]	1.45[.66]	6.73 [3.34***]	4.25 [1.14]	1.81 [1.81*]	2.08 [.95]
Adj. R-squared	.12	.23	.23	.30	.10	.08	.09	.05

Statistical significance: *90% **95% ***99%

4.2.2 Use of performance information

Tables 7 and 8 report regression results that assess the correlation between the quality of an agency's GPRA initiatives and uses of performance information. Similar to table 4, table 7 estimates the correlation directly, regressing the 2007–2000 difference in various uses of performance information on the 2007–2000 difference in Score and Leadership. Similar to table 5, table 8 estimates the relationship between the 2007–2000 difference in the availability of outcome and output measures and the 2007–2000 difference in uses of performance information.

The results in both tables indicate that better GPRA reports are associated with greater use of performance information for certain purposes. In table 7, the *Scorecard* variables have a statistically significant correlation with the use of performance information in allocating resources, developing or refining program performance measures, and setting or revising program goals. Thus, table 7 suggests that managers in agencies with better

[14] To conserve space, regression results are reported in this section only when the Score variable was statistically significant at the 90 percent level or greater.

GPRA reports are generally more likely to use performance information to allocate resources, devise measures, and set goals.

Table 8 shows that availability of outcome and output measures has statistically significant correlations with use of performance information to allocate resources, set priorities, change programs or work processes, and set goals. Score may be indirectly related to these uses of performance information via its correlation with outcome and output measures.[15] Including an output measure variable in table 8's regressions, however, usually reduces the statistical significance of the outcome variable even though the output variable is rarely statistically significant.

The one exception is for allocating resources. Improvements in the availability of output measures have a negative and marginally significant correlation with use of performance information to allocate resources. This is also the only specification in which the *Scorecard* score is marginally significant and positive. The positive direct effect of the *Scorecard* variable outweighs the negative indirect effects via the output variable.[16] These coefficients may be statistical flukes, or they may be a cautionary tale about excessive focus on output measures to the exclusion of outcome measures.

Table 7: Difference in GPRA report quality and difference in uses of performance information

22 observations	**Allocate**	**Allocate**	**Measures**	**Measures**	**Goals**	**Goals**
Score	.54 [2.25**]	.58 [2.51**]	.58 [2.44**]	.61 [2.58**]	.52 [1.95*]	.60 [2.50**]
Leadership		.30 [1.62]		.22 [1.15]		.501 [2.59**]
Constant	3.30 [1.61]	-1.64 [-.45]	5.61 [2.76**]	2.02 [.54]	6.10 [2.65**]	-2.03 [-.54]
Adj. R-squared	.16	.23	.19	.28	.16	.31

Statistical significance: *90% **95% ***99%

[15] Several alternative specifications that are not reported led to essentially the same results. When the outcome variable is used as the only explanatory variable, it is highly statistically significant. When the leadership and validity control variables were included, they were not statistically significant and did not materially alter the results.

[16] Table 6 indicates that a 1-point increase in *Scorecard* score or rank generates about a 0.4 percentage point increase in the incidence of output measures. Multiplying 0.4 times -0.45 (the output coefficient in table 8) yields -0.18. This is outweighed by the positive coefficient of approximately 0.3 on *Scorecard* score/rank in table 8.

Table 8: Difference in existence of measures and uses of performance information

22 observations	**Allocate**	**Allocate**	**Priorities**	**Priorities**	**Change**	**Change**	**Goals**	**Goals**
Outcome	.77 [5.05***]	.86 [5.66***]	.41 [2.29**]	.36 [1.89*]	.55 [2.09**]	.48 [1.69]	.50 [2.12**]	.48 [1.89*]
Output		-.45 [-1.83*]		.24 [.76]		.37 [.81]		.08 [.19]
Score	.18 [1.00]	.31 [1.71]	.13 [.61]	.06 [.24]	-.23 [-.76]	-.34 [-1.01]	.29 [1.07]	.27 [.88]
Constant	-2.34 [-1.32]	-.91 [-.50]	3.04 [1.46]	2.29 [.99]	4.10 [1.33]	2.93 [.85]	2.45 [.90]	2.21 [.71]
Adj. R-squared	.66	.67	.22	.21	.10	.08	.25	.21

Statistical significance: *90% **95% ***99%

4.3 Causality

In both the pooled regressions and the difference regressions, the relationship between the quality of an agency's GPRA initiatives and managers' possession and use of performance information appears pretty robust. The positive relationship is consistent with the classic theories of bureaucratic behavior developed by Tullock and Downs. GPRA articulated what elected leaders expected agencies to do. The Office of Management and Budget continued to make GPRA performance measurement and reporting a priority. This, plus the existence of the law itself, was apparently a strong enough signal to prompt some progress. GPRA drove agencies to develop and use performance measures, and this focus filtered down through the managerial hierarchy to affect performance measurement at levels below that of the agency-wide goals and measures.

However, causality might also run in the other direction. Perhaps other factors unrelated to GPRA drove agencies to develop and implement performance measures. The agencies that are better at this would naturally produce better GPRA reports as well.

This second interpretation also explains the positive correlation between the quality of agency GPRA reports and GAO survey responses. But it is less plausible than the explanation offered in this paper. GPRA was enacted in 1993. It thus predates the earliest GAO survey results used in this paper by seven years. To believe that agency performance management initiatives undertaken for other reasons drove most of the improvements in GPRA reporting, one would have to believe that agency managers largely ignored GPRA, but then developed and implemented many of the types of performance measures GPRA mandated anyway. This interpretation relies heavily on coincidence to explain the correlation between GPRA report quality and GAO survey results.

Even if this second interpretation is true, it still posits a positive correlation between GPRA report quality and GAO survey responses relevant to performance measurement. This suggests that, regardless of the direction of causality, GPRA is complementary to, rather than a distraction from, agency performance measurement initiatives.

5. How big was the effect?

Statistical significance is not quantitative significance (McCloskey and Ziliak 1996). The results in section 4 indicate that the quality of an agency's GPRA initiatives have a positive and significant correlation with the availability and use of performance information. But is this effect large or small?

Table 9 offers one answer. The table multiplies the coefficient on the *Scorecard* variable from tables 3–5 by 34, the average *Scorecard* score. This shows how a report with an average score would have affected the percentage of managers saying they have or use performance information. The third column of table 9 shows the actual average for each survey question for comparison.

Separate figures are calculated for 2000 and 2007 since, in some cases, the Year*score variable indicates that the effect of the *Scorecard* score differs in the two years. Table 9 indicates that the average *Scorecard* score of 34 would account for a noticeable but not huge fraction of the positive responses to the GAO survey questions on the availability and use of performance information.

In both years, for example, the *Scorecard* score appears to account for about 10 percentage points of the positive response to the question about the availability of outcome measures. The actual average positive response was 50.4 percent. Thus, about one-fifth of the incidence of outcome measures could be attributed to the quality of the "average" agency's GPRA initiatives. In most other cases where direct effects are estimated, the average score is responsible for about one-quarter of the observed positive responses in 2000 and 15–20 percent in 2007. Proportions are somewhat lower for the indirect effects in the lower part of the table.

Of course, a higher *Scorecard* score could lead to an even larger proportion of positive responses. The highest score earned in this data set, for example, was 55. A score of 55 would generate a 16.5 percentage point increase in the number of managers saying they have outcome measures for their programs. This equals one-third of the actual average of 50.4 percent.

Table 9: Effects of average Scorecard score of 34

	2000	2007	Actual average
Existence of performance measures			
Outcome	10.2	10.2	50.4
Output	16.0	7.5	57.1
Efficiency	9.9	9.9	41.1
Use of performance information			
Allocate resources	11.9	7.1	48.7
Set priorities	13.6	8.5	50.2
Coordinate with other parties	9.2	9.5	40.9
Dev./refine measures	11.6	7.8	41.8
Set/revise goals	11.6	7.1	47.3
Indirect effects on use of performance information			
Allocate resources	7.0	7.0	48.7
Set priorities	12.4	7.7	50.1
Change processes	9.3	6.3	47.5
Coordinate with other parties	7.4	5.2	40.9
Dev./refine measures	7.1	7.1	41.8
Set/revise goals	6.5	6.5	47.3
Set job expectations	7.2	3.4	49.3
Reward employees	4.7	4.7	48.7
Manage contracts	6.2	6.2	26.6

The results of the regressions using differences in tables 6–8 also suggest the *Scorecard* score has a noticeable effect on the availability and use of performance information. Table 10 calculates the effect of several postulated changes in *Scorecard* scores, using the most statistically significant *Scorecard* coefficients from the difference regressions in tables 6–8.

Not surprisingly, an agency that achieved the lowest change in *Scorecard* score in the sample—1 point—would show little change in the availability and use of performance information attributable to GPRA. The average change, however, was approximately 7 points.[17] A 7-point change in the *Scorecard* score is associated with a noticeable change in the percentage of managers who have or use performance information. This change often equals one-quarter to two-thirds of the actual average change calculated in table 2.

[17] Some agencies saw improvements in their scores, while others saw declines. The average size of the absolute value of the change—whether positive or negative—was 6.95 points. This is different from the much lower mean value reported in appendix I. Positive and negative changes tend to cancel each other out, leaving a mean close to zero.

Table 10: Effects of Scorecard point changes

	Smallest score Change (1 point)	**Average Score Change (7 points)**	**Largest score change (17 points)**	***Actual Average Change (Table 2)***
Existence of performance measures (Table 6)				
Outcome	0.5	3.6	8.8	8.4
Output	0.4	2.9	7.1	6
Customer satisfaction	0.5	3.2	7.7	6.6
Any	0.3	2.0	4.8	2.6
Use of performance information (Table 7)				
Allocating resources	0.6	4.1	9.9	4.9
Dev./refine measures	0.6	4.3	10.4	6.3
Set/revise goals	0.6	4.2	10.2	6.9
Indirect effects on use of performance information (Table 8)				
Allocating resources	0.6	4.0	9.7	4.9
Setting priorities	0.2	1.5	3.7	6.7
New approach/process	0.3	2.0	4.9	8.8
Set/revise goals	0.3	1.8	4.4	6.9

Especially noteworthy is the effect on one key use of performance information GPRA sought to encourage: resource allocation. Using the directly estimated coefficient from Table 7, the 7-point average *Scorecard* score change is associated with a change in the use of performance information to allocate resources that is almost as large as the average change that actually occurred. Using the indirect estimate from table 8 generates a similar change.

Finally, the largest score change observed (17 points) is associated with changes in the availability and use of performance information that are often much larger than the actual average change.

These calculated changes probably under-estimate the total effect of GPRA on the availability and use of performance information. They estimate the possible *marginal* effects of changes in the quality of agencies' GPRA initiatives on the availability and use of performance information. This is not the same as estimating how the GAO survey results would look with and without GPRA. The data are from 2000 and 2007; GPRA was enacted in 1993. Any effects of GPRA that had already occurred before 2000 or do not vary with the quality of an agency's GPRA report are not captured in this paper's regression equations or calculations.

Table 9 and, especially, table 10 suggest that the quality of agencies' GPRA initiatives between 2000 and 2007 is associated with large changes in the availability and use of performance information documented in the GAO surveys. In that sense, the GPRA-performance measurement relationship is quantitatively significant.

Whether the total measured improvement should be considered large or small depends on one's frame of reference. Actual improvements in the availability and use of performance information reported in table 2 rarely averaged more than 10 percentage points total over a seven-year period, which translates into an improvement rate of slightly more than 1 percent annually. This does not sound impressive. Likewise, it is not clear whether the percentages of affirmative responses for 2007 should be considered acceptable or disappointing 14 years after GPRA's enactment and nine years after agencies started producing annual performance reports. Fewer than two-thirds of managers report that they measure outputs to a great or very great extent. Only about half report that they measure outcomes to a great or very great extent. The survey results on use of GPRA strategic goals are more encouraging, with about three-quarters of managers reporting that they use their agency's strategic goals to a great or very great extent for various purposes. Yet only about half of managers at most say that they use performance information for the specific purposes enumerated by GAO.

Compared to the theoretical ideal—which might be defined as close to 100 percent positive responses for all questions—the 2007 results and the improvements associated with GPRA look somewhat disappointing. However, it is doubtful that the ideal is actually achievable. Radin (2006) points out numerous practical difficulties that make measurement of outcomes for all programs difficult or impossible. Tullock (2005) and Downs (1962, 136–43) present compelling logic that any bureaucracy will suffer from significant inefficiencies when compared to the theoretical ideal in which every employee of the organization spends every minute doing precisely what the ultimate "sovereign" (the voters) would want each one to do. Nevertheless, a bureaucracy may be the most efficient method available for accomplishing some tasks, when compared to reasonably achievable alternatives (Niskanen 1974, 191; Mises 1946, 48–49). Viewed in this light, these statistics need not indicate a deficiency.

6. Conclusion

The theory and evidence in this paper suggest that GPRA has been a significant factor encouraging the development and use of performance information in major federal agencies. Since there were many good reasons to believe GPRA would have little impact, it's comforting to see evidence suggesting that GPRA truly has had a systemic effect. GPRA has made performance information more available, and managers have used that information to make decisions.

Yet progress has been uneven across agencies. Improvements in the availability and use of performance information have been uneven because, as GAO predicted in 1997 (GAO 1997), GPRA implementation has been uneven. If the quality of an agency's GPRA report indicates the quality of its GPRA implementation, the Mercatus *Scorecard* and evaluations by the Association of Government Accountants both show that some agencies have done much better than others. The rate of improvement in GPRA reporting has also been uneven across agencies. Since better GPRA reporting is correlated with more available and more widely used performance information, it should be no surprise that agencies show variation in the availability and use of such information.

These findings suggest several avenues for further research:

(1) Leadership has a positive and very statistically significant correlation with the availability and use of performance information. Have all agencies achieved the extent of GPRA implementation that is realistically achievable, given the constraints? Or are there still achievable opportunities for improvement that could be promoted by innovative leadership?

(2) A higher proportion of the agency budget devoted to regulation or research and development is sometimes correlated with less availability or use of performance information. One barrier to improvement may be the types of constraints identified by Radin (2006), which make performance measurement and management more difficult and/or costly for certain types of programs. If so, are there opportunities either to reduce these difficulties or to modify the GPRA framework to accommodate agencies or programs for whom performance measurement is unusually difficult?

(3) A barrier to across-the-board improvement may be mis-alignment of incentives, as principal-agent theory would suggest. If so, how can elected leaders better alter public managers' incentives to promote performance management?

(4) A final barrier to generalized improvement could be the incentives faced by elected leaders themselves. If so, how might elected leaders be made more responsive to the wishes of the general public or the median voter interested in performance, rather than parochial interests concerned about receiving expenditures?

(5) Has the actual performance of the federal government improved as a result of the increased availability and use of performance information reported by federal managers? If so, then questions 1–4 above should be important to public managers and policymakers as well as scholars.

References

Brito, Jerry, and Jerry Ellig. Forthcoming. "Toward a More Perfect Union: Regulatory Analysis and Performance Management," *Florida State University Business Review* 8.

Clinton, Joshua D., and David E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences," *Political Analysis* 16: 3-20.

Curristine, Teresa. 2002. "Reforming the U.S. Department of Transportation: Challenges and Opportunities of the Government performance and Results Act for Federal-State Relations," *Publius* 32: 25-44.

Downs, Anthony. 1967. *Inside Bureaucracy*. Boston: Little, Brown and Co.

Frisco, Velda, and Odd. J. Stalebrink. (undated manuscript) "Congressional Use of the Program Assessment Rating Tool."

Gilmour, John B., and David E. Lewis. 2006. "Does Performance Budgeting Work? An Examination of the Office of Management & Budget's PART Scores," *Public Administration Review* (Sept./Oct.):

Government Accountability Office. 2005. *Managing for Results: Enhancing the Use of Agency Performance Information for Management Decision Making*, GAO-05-927 (September).

_____. 1997. *The Government Performance and Results Act: 1997 Governmentwide Implementation will be Uneven*, GAO/GCD-97-109 (June).

Julnes, Patria de Lancer, and Marc Holzer. 2001. "Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementaiton," *Public Administration Review* 61 (Nov.-Dec.): 693-708.

Kettl, Donald F. 1997. "The Global Revolution in Public Management: Driving Themes, Missing Links," *Journal of Policy Analysis and Management* 16 (Summer): 446-62.

McCloskey, Deridre N., and Stephen T. Ziliak. 1996. "The Standard Error of Regressions," *Journal of Economic Literature* 34 (March): 97-114.

McNab, Robert M., and Francois Melese. 2003. "Implementing the GPRA: Examining the Prospects for Performance Budgeting in the Federal Government," *Public Budgeting & Finance* 23 (Summer): 73-95.

McTigue, Maurice, Jerry Ellig, and Steve Richardson. 2001. *2nd Annual Performance Report Scorecard*. Arlington, VA: Mercatus Center, George Mason University.

McTigue, Maurice, Henry Wray, and Jerry Ellig. 2008. *9th Annual Performance Report Scorecard.* Arlington, VA: Mercatus Center, George Mason University.

Mihm, J. Christopher, and James R. White. 1999. *Elements of Successful Improvement Initiatives, Testimony*, GAO-T/GGD-00-26 (October 15).

von Mises, Ludwig. 1946. *Bureaucracy*. New Haven: Yale University Press, 1946.

Niskanen, William A., Jr. 1994. *Bureaucracy and Public Economics*. Brookfield, VT: Edward Elgar.

Norcross, Eileen, and Joseph Adamson. 2007. "An Analysis of the Office of Management & Budget's Program Assessment Rating Tool (PART) for Fiscal Year 2008," Working Paper, Mercatus Center at George Mason University.

Radin, Beryl A. 2006. *Challenging the Performance Movement*. Washington, DC: Georgetown University Press.

Han Chun, Young, and Hal G. Rainey. 2005a. "Goal Ambiguity in US Federal Agencies," *Journal of Public Administration Research and Theory* 15: 1-30

_____. 2005b. "Goal Ambiguity and Organizational Performance in US Federal Agencies," *Journal of Public Administration Research and Theory* 15: 529-57.

Roberts, Nancy. 2000. "The Synoptic Model of Strategic Planning and the GPRA: Lacking a Good Fit With the Political Context," *Public Productivity & Management Review* 23 (March): 297-311.

Scott, Graham, Ian Ball, and Tony Dale. 1997. "New Zealand's Public Sector Management Reform: Implications for the United States," *Journal of Policy Analysis and Management* 16 (Summer): 357-81.

Steinhardt, Bernice. 2008. *Government Performance: Lessons Learned for the Next Administration on Using Performance Information to Improve Results*, GAO-08-1026T (July).

Tullock, Gordon. 2005 [1965]. *The Politics of Bureaucracy*, reprinted in Charles K. Rowley, Ed., *The Selected Works of Gordon Tullock*. Indianapolis, IN: Liberty Fund.

Appendix I: Summary Statistics on Data

Variable	Obs	Mean	Std. Dev.	Min	Max
Scorecard variables					
Score	46	33.80435	8.261341	17	55
Year*score	46	18.04348	18.5004	0	55
Availability of performance measures					
Outcome	46	50.38696	10.73291	30.1	71.7
Output	46	57.10217	10.38382	39.6	85.2
Efficiency	46	41.11087	10.30711	21.1	67.4
Customer satisfaction	46	40.01087	12.17891	13.6	74.2
Quality	46	40.39783	11.24423	23.5	67.5
Any	46	88.8587	5.175609	74.2	99
Uses of performance information					
Allocate resources	46	48.72609	8.370329	30.1	65.9
Set priorities	46	50.16087	9.958435	25.9	69.3
New processes	46	47.54348	10.14908	27	63.9
Coordinate with external orgs.	46	40.9113	9.340189	19.82	57.5
Develop or revise measures	46	41.84565	9.432514	22.6	60.5
Set or revise goals	46	47.25	9.487389	28.1	66.6
Establish job expectations	46	49.28043	10.80842	21.6	69.6
Reward employees	46	48.67826	10.61636	23.6	70.9
Manage contracts	46	26.61087	10.72454	7.9	54.4
Control variables					
Leadership	46	63.47826	12.71326	29.8	89.6
% Competitive grant	46	8.56304	18.99115	0	76.7
% Block/formula grant	46	16.27609	23.39782	0	77.5
% Regulatory	46	6.7087	20.52648	0	100
% R&D	46	9.37609	22.49635	0	100
Agency received CEAR	46	.3043478	.4652151	0	1
Clintonlewis (agency ideology)	46	.023913	.9308192	-1.43	2.21
Year 2007	46	.5217391	.505047	0	1
Outlays	46	98.78328	178.2812	.031	672

All variables below are 2007-2000 differences unless otherwise noted.

Variable	Obs	Mean	Std. Dev.	Min	Max
Scorecard variables					
Score	22	1.590909	8.633595	-14	17
Rank	22	.4090909	8.617032	-17	15
Uses of performance information					
Set program priorities	22	6.554545	8.487623	-12.8	24.7
Allocate resources	22	4.154545	10.34287	-12.8	35.8
New processes	22	8.195455	11.65102	-17.1	35.9
Coordinate with external orgs.	22	9.758182	7.361715	-5.4	20.6
Develop or revise measures	22	6.531818	10.42129	-20.5	27.1
Set or revise goals	22	6.936364	11.30868	-14.1	25.1
Establish job expectations	22	12.85909	10.99479	-9.3	33.9
Reward employees	22	9.027273	13.37785	-19.8	31.1
Manage contracts	22	5.104546	8.651313	-10.4	28.8
Availability of performance measures					
Any (% answering yes or no)	22	2.186364	5.625393	-9.1	15.1
Output	22	5.295455	6.604506	-8.3	16.8
Efficiency	22	9.418182	5.458906	.8	20.5
Customer satisfaction	22	7.404546	9.79526	-7.1	33.2
Quality	22	6.454546	7.717743	-5.1	20.1
Outcome	22	8.072727	9.956725	-6.3	39
Control variables					
Leadership commitment	22	16.00909	10.64889	-3.8	36.3
Validity of measures	22	7.654545	8.629944	-3.2	24.2
Difference outlays 2007-2000	22	32.57273	63.70314	-2.7	222.4