

FOUR GUIDING PRINCIPLES FOR IMPROVING FEDERAL PERFORMANCE MANAGEMENT

Federal programs have made noteworthy progress since passage of GPRA using outcome-focused goals and measures to improve program impact, cost-effectiveness, and accountability, but much work still needs to be done. To stimulate needed innovation and discovery and to achieve the performance improvement breakthroughs that should be possible, the following four principles, which address the performance-inhibiting problems described in the Findings section of this report, should guide future federal performance management improvement efforts.

Principle One: Communicate Performance Trends and Targets, Not Target Attainment and Ratings

Government should pay increased attention to communicating performance trends and targets. Despite the fact that federal agencies have been expected to set agency goals, measure progress toward them, and report performance to the public for more than ten years under the Government Performance and Results Act of 1993 (GPRA) and to do the same at the program level since the inception of PART, it is still surprisingly difficult to find federal performance trend information. OMB and agency summary charts have not shown performance trends.

Instead, federal performance reports and summary charts show the number and percentage of “targets attained,” PART ratings, and traffic light scores on the President’s Management Scorecard as primary indicators of program and agency performance. This has focused agencies on meeting targets and raising PART and President’s Management Scorecard ratings but diverted their attention from undertaking the full range of assessments, analyses, and actions needed to improve societal conditions.

OMB reinforced the message to agencies, perhaps inadvertently, that performance trends were not as important as target attainment and PART ratings when it chose as exemplary agency Performance and Assessment reports (in its 2008 A-11 budget instructions) the reports of four agencies that showed the percentage of targets met, not the direction and size of performance change, to summarize agency performance.

Targets are powerful management tools, especially when they specify factors such as time, quantity, place, and population, etc. They are useful for focusing, motivating, and communicating priorities within an organization and to people beyond it and for enlisting outside assistance and resources. PART reviews, too, provide useful feedback to agencies on areas of program strength and weakness. The percentage of targets attained and PART performance ratings do not, however, effectively or objectively communicate performance.

It is far more informative and objective to communicate whether, where, in what direction, and by how much performance and related indicators are moving. Reporting performance trends indicates whether or not program outcomes and interim outcomes are going in the direction desired, suggesting whether agency actions are working as intended, not simply whether a target has been met or a commitment fulfilled. Reporting trends also highlights sudden or unexpected changes in direction and size. When agencies follow up on these unexpected changes to understand their underlying reasons, it often leads to the discovery of effective government interventions worthy of replication. It can also lead to the discovery of underlying causal factors contributing to performance declines (or gains) that government can influence.

“Target Attainment” is a Limited Indicator of Performance

Targets are a powerful management and communication tool, but reporting target attainment communicates control, not priorities, problems, and progress. Measuring the percentage of targets met conveys little information about performance to those not involved in target negotiations unless the targets are known, the reasons they were selected are understood, and the reasons and targets themselves are accepted. The emphasis on target attainment sends a message that Congress and the public should accept the targets selected by agencies with OMB approval.

Measuring the number or percentage of “targets attained” can be a useful indicator for internal management purposes, but “targets attained” is not a good performance indicator for multiple reasons:

- Knowledge of targets is essential to the value of “target attainment” as a performance indicator.** While it has been relatively easy to find summary charts of federal performance showing the number and/or percentage of targets attained, PART scores, and PART ratings, it is not easy to find a summary list, even by agency, of the targets set by federal programs. Interested readers must dig down into the details of agency and program performance reports to identify what the targets are. Summary charts that show target attainment but fail to show what the targets are, which were exceeded, which were attained, and which were not attained communicate only whether agencies met White House and agency leaders’ expectations, but not actual government performance. Without knowledge of what the targets are, summary charts reporting that a program or agency attained, say, 85 percent of its targets convey little useful information about the state of the world and program performance.
- Acceptance of targets is essential to the value of “target attainment” as a performance indicator.** For Congress and others to care about target attainment rates, they must not only know, but also accept, the reasonableness of the targets chosen. While Congress sometimes writes targets in law that agencies then use when setting targets, agencies more often must narrow the broad goals and even the specific targets

Congress sets because budgetary resources are insufficient to tackle them all. Explicitly or implicitly, agencies make decisions about which problems, which populations, and which geographic areas to serve first. To understand these priority-setting decisions, Congress and the public need to know not only what the specific targets are, but also the reasons agencies choose them. It can be hard to find these reasons in agency documents.

- Reporting on target attainment is not helpful unless agencies confirm awareness and acceptance of the targets with key congressional offices.** Congressional staff interviewed for this report indicated that they heard little from agencies about what their performance data showed, why agencies selected the targets they chose, why they chose specific strategies, and where fund reallocations among targets might be appropriate.

With a few noteworthy exceptions, most exchanges that did occur were formal, and confined primarily to official documents and congressional hearings. This seldom afforded a useful means for clarifying misunderstandings and digging more deeply into unanswered questions. Some congressional appropriations staff interviewed for this report indicated they would welcome more informal discussions about targets, trends, and strategies, although agency and OMB officials indicated that some agencies offered to brief key congressional committees with little response.
- Target attainment does not always correlate with performance gains.** An agency or program that chose targets lower than prior-year performance could meet all of its targets even if actual performance had declined. Adoption of more lenient targets did occur, yet neither agencies nor OMB tracked the number or percentage of targets set at levels lower than prior-year performance and target levels. Without companion information showing how targets compared to past performance, summary charts indicating targets met or even exceeded cannot accurately convey the direction of performance change.
- Tracking target attainment rather than performance trends as the primary performance indi-**

cator can motivate measurement manipulation.

Experience shows that some organizations, eager to meet targets or earn higher ratings, “cream-skim.” They serve populations or places more easily served. “Cream-skimming” can also arise when agencies track trends, but the intensified pressure of meeting a target makes it more likely to occur when target attainment is used as the primary performance indicator.

- **Tracking target attainment also intensifies the temptation to choose timid targets that programs know they can meet, rather than the sort of “stretch” targets that research has shown can lead to larger performance gain.** To earn high PART scores, programs were expected not only to set ambitious long- and short-term targets, but also to meet them. By definition, this is highly unlikely. Ambitious targets are those that cannot be met all of the time. In other words, the chance that a program could earn a “yes” score on the two PART questions about ambitious targets and also on the PART question about meeting targets is, by definition, extremely small. The PART scoring penalty for programs that failed to meet their targets was exacerbated by the summary charts that implied, by their use of target attainment percentages, PART scores, and PART ratings, that programs that dared to set but failed to meet ambitious targets had lower performance.

Ratings are Limited Indicators of Performance

OMB and the White House used PART performance ratings (effective, moderately effective, adequate, ineffective, and results not demonstrated) as another performance indicator. Using PART ratings (rather than performance trends) as a performance indicator was a sensible start-up strategy. It provided a mechanism for recognizing progress while programs built their capacity to measure outcomes. However, using PART ratings as an ongoing way to measure and communicate program performance, or the rating of the President’s Management Agenda Scorecard, has serious limits for the following reasons:

- **Ratings are subjective and therefore likely to be inconsistent and reflective of reviewer bias.** The structure of the PART review process that produces PART ratings, with single but different individuals making multiple decisions about the

adequacy of program practices along multiple dimensions, creates inconsistency and bias problems.

Inconsistent PART reviews are not the fault of individual OMB examiners; they are inherent to any review dependent on the opinion of an individual evaluator. Olympic scoring for events such as gymnastics or diving, where winners are determined not by the cross of a finish line but by opinion, accepts the inevitability of reviewer bias. To contain it, expert judges from different countries use explicit scoring criteria. Even then, reviewer bias is anticipated, and the high and low scores are tossed out. With all these adjustments, bias by judges is still seen as a problem.

Reviewer subjectivity creates problems beyond inconsistency. As the vast literature on cognitive bias has found, reviewers are influenced by their professional perspectives, experience, expertise, and values. OMB examiners, by the nature of their budgeting and central office responsibilities, are likely to hold a professional bias toward cost-cutting over performance improvement. They are also likely to have an inclination to play more of a controlling than an assisting role.

- **Ratings hide multiple, relevant dimensions of performance.** Ratings, whether grades or labels such as “effective” or “ineffective,” hide valuable information when distinct dimensions of product performance are not easy to see. Consider how *Consumer Reports (CR)* presents its ratings to make them useful to readers. *CR* charts show how each product fares on multiple dimensions of performance. *CR* communicates its summary preferences in two ways: ranking in order of quality across all performance criteria and highlighting “best buys” that, while not the top quality product, are considered best for the price. Summary charts also show how each product fares on each performance dimension, allowing readers to assess whether they agree with *CR*’s summary assessment and to adjust that assessment to incorporate their own preferences.

OMB summary charts of PART ratings do not make it easy for readers to see how well a program fared on different dimensions of performance. Summary charts showing performance

trends for each key indicator would convey that information more succinctly.

- **Ratings influence choices among alternative products and qualify entities for certain situations, but PART ratings are not needed for those purposes.** *Consumer Reports* ratings help buyers compare products. Ratings are also useful for qualifying people and organizations to enter specific competitions, assume responsibilities, or earn special privileges. The use of a PART rating is less clear, however. Neither Congress nor OMB is likely to use PART ratings to choose among producers of different federal functions, because federal programs seldom compete directly with one another; they tend to be created to address very distinct needs not provided by any other program. Programs with similar functions can find it helpful to look at other federal programs with a higher PART rating to identify practices worth replicating, but they could gain more insight by looking at scores on individual PART questions and changes in the scores over time.

Ratings, arguably, motivate some programs, but not necessarily in the ways intended. For example, several agency officials report that, over time, they learned how to “get to green” on the President’s Management Agenda Scorecard and improve their PART score by figuring out minimal actions they could take to earn an upgrade even though the changes did not improve operational or program effectiveness. Actual performance trends are likely to be a better indicator of progress than ratings and are more likely to align agency actions with organizational objectives.

- **Ratings do not recognize continuous improvement.** The PART rating system cannot recognize improvement in areas where an agency has already earned the highest “effective” score. OMB could add new conditions to motivate further improvement, but that would have the unfortunate side effect of lowering scores for those who had already reached the top. Ratings have no way to recognize further improvement by those already doing well.

Principle Two: Encourage Performance Improvement with Increased Diagnostic Analysis, Data-Driven Discussion, Practical Experiments, and Knowledge Sharing

Principle One argued that agencies and OMB should communicate targets and trends when reporting agency and program performance. Reporting is not, however, measurement’s only or, arguably, its most important use. Measurement is most valuable when organizations use the data they collect not just to report, but to illuminate, communicate, motivate, and allocate.

Recent federal performance management practices have paid too little attention to the diagnostic analysis of data to understand the nature of problems more precisely, the factors affecting performance trends and variations, and the government actions that can influence them. They also paid too little attention to sharing insights from those analyses with people in government and others that design and deliver programs.

Agencies need to pay more attention to:

- Understanding the size and characteristics of problems to be addressed and opportunities to be pursued
- Discovering why performance levels change or vary
- Finding effective interventions for different types of problems
- Sharing knowledge of problems and solutions so it can be applied in program implementation

Agencies also need to pay more attention to sharing the data they collect to make it easier for others to analyze, to discover patterns and possibilities, and to inform policy decisions and personal choices.

Specifically, federal agencies and programs need to use the performance measurement they collect to:

- **Illuminate problems needing attention, their causes, and interventions worthy of replication.** Analyzing performance data helps agencies determine the size and characteristics of social

and system problems needing attention so they can set priorities among them. It reveals promising programs worth continuing and problems that need adjustment. Studying unexpected changes, variations in performance levels, and anomalies can lead to a deeper understanding of causal factors that can be influenced to improve outcomes.

- **Communicate measurements, lessons and data.** Communicating and discussing performance trends for priority targets sends the message that previously set targets continue to be a priority. Communicating performance trends, supported by information about the timing of new interventions, helps to speed adoption across agency delivery partners of practices that improve trends and slow adoption of those that do not. Communicating data helps to stimulate external analysis and support coordination among multiple delivery partners cooperating to advance shared goals.
- **Motivate with measurement by using it to provide fast feedback.** This can energize people by providing a sense of accomplishment when progress is being made and by creating a sense of urgency for making programmatic changes when it is not.
- **Allocate resources to activities with the greatest performance impact relative to time, money, and other resources invested.**

When analysis of performance data does not reveal effective, efficient practices, federal agencies may need to experiment to discover increasingly effective and cost-effective interventions. Once promising practices are found, they may also need to experiment to find successful methods for promoting their adoption by other agencies and delivery partners.

Principle Three: Present Information To Meet the Needs of Specific Audiences

One of the most significant findings from the interviews conducted for this study is that, despite near consensus about the value of performance measurement, so few—in Congress, in agency field and headquarters offices, among those served or regulated, in advocacy organizations—found the

numerous documents and websites with federal performance information useful. There were some noteworthy exceptions, such as the Department of Education’s budget justification, which was singled out as a model for other agencies in a House appropriations bill. Most performance-linked documents were viewed as confusing and disappointing, however. They paid too little attention to figuring out who wanted and needed performance information, how they could use it, and how to meet those needs.

There are several possible explanations for this problem. Budget justifications, the content of which was specified by OMB Circular A-11, irritated congressional appropriators because current year budget proposals could not be compared to prior-year information. The GPRA-required performance reports that agencies submitted to Congress were overwhelming because they packaged copious amounts of financial and management risk information together with performance data in a single Performance and Accountability Report (PAR), the content of which was prescribed by OMB Circular A-136. The emphasis on target attainment status and commitment fulfillment rather than goals, targets, trends, analysis, and strategies rendered most performance reports of limited value to would-be users. Also, while the President’s Management Agenda Scorecard ratings functioned as a useful checklist of management activity expectations for the White House, it imparted little information to others.

In addition, as previously noted, performance trend information, of great interest to the public and those looking for successful programs to replicate, was remarkably hard to find. Too often, PART reviews and agency performance reports shared only a few years of performance trend data and only for some, but not for all, of the relevant indicators. When agencies

The Reports Consolidation Act of 2000

The Reports Consolidation Act of 2000 (P.L. 106-531) authorized agencies to consolidate their financial and performance reports, perhaps encouraging but not mandating the consolidation that produced the PAR reports everyone found so overwhelming. OMB Circular A-136 translated that authority to a requirement.

The Program Assessment Rating Tool (PART)

PART was launched as a pilot in February 2002 and introduced in the President's FY 2003 budget as a tool to be applied to every government program. PART is a list of "Yes/No" questions asking programs how they use goals, measures, and evaluation. The questions are grouped in four categories:

- Program purpose and design
- Strategic planning
- Program management
- Program results/accountability

(Appendix I contains the full list of PART questions.) PART is more than a list of questions, though. It also involves third-party review and public reporting.

PART tackled what agencies and Congress identified as one of GPRA's biggest weaknesses—the use of what some described as the "10,000-mile high indicators" used in GPRA documents. While GPRA goals and measures could be powerful when used by senior management to drive change, enhance coherence, and spur cooperation across agency programs, they often lacked relevance for program offices and congressional decision-makers unless agencies explicitly articulated what was expected of each program to advance GPRA goals. Some agencies took the time to sort out and communicate the link between GPRA goals and program expectations—what some call cascading down and rolling back up. Most, however, did not.

PART translated the goal-setting and measurement requirements of GPRA to the program level, the scale at which most agency operations function and at which funding decisions are made. By shifting attention to programs and at the same time announcing that OMB would review 20 percent of agency programs each year so all programs were reviewed at least once every five years, PART signaled that all federal programs, not just agency central-office staff, were expected to adopt outcome-

focused performance management practices. Programs could request follow-up PART reviews (before the scheduled five-year review) to earn a higher PART score and rating earlier, and many did.

PART uses several distinct motivational mechanisms to increase agency attention to goals and measurement:

- **Focus on the Program Level.** PART requires programs, not agencies, to select goals and measures. This pushed responsibility for goal-setting and measurement adoption beyond the central budget, planning, or performance office of an agency to the program level.
- **Questions.** PART provides a detailed set of questions that every federal program is required to answer at least once every five years. This compelled agencies to consider basic program management questions that program managers often wanted but never found the time to consider.
- **External Review.** OMB, not the agency, conducts the reviews of each program's performance management practices, which compels agencies to consider how each program fares relative to each of the PART questions. Programs cannot ignore a question as irrelevant without being able to defend that decision to OMB. The review process varies by OMB reviewers and is up to the discretion of the examiner.
- **Scores.** OMB provides a binary "Yes/No" score and commentary on each question, providing programs with feedback about specific areas of strength and weakness. Questions are weighted and summed to tally a total program score, up to 100 percent. The PART score for each question is posted for public review on the Internet, along with short explanatory notes. The aggregate PART score is also posted.
- **Ratings.** OMB uses the total PART score to determine a rating: effective, moderately effective, adequate, ineffective, and results not demonstrated. OMB rates a program "results not dem-

onstrated” when it concludes that insufficient measurement is available to determine program impact. Some agencies reported their programs were motivated to earn a higher rating.

- **Improvement Commitments.** Following a PART review, each agency and OMB agree on a list of specific actions each program commits to take. Programs periodically provide OMB and the public, via the web, a written update on actions they have taken to fulfill the commitments.
- **Transparency.** Program ratings are prominently displayed in OMB summary charts and OMB shares the following underlying scoring details with the public:
 - The PART score for each question
 - Comments on each question
 - The total score for each program
 - An overall program rating
 - The list and status of improvement commitments

All of this information was available to view at www.expectmore.gov, the OMB-run website which facilitated public access to PART reviews, strategic plans, and performance reports.

In addition, a second OMB website, www.results.gov, contained hints and examples intended to help agencies improve their performance management practices. This site also featured summary data about the status of agency implementation of the President’s Management Agenda Scorecard.

were required to submit other annual reports to Congress with relevant performance and other data, too few agencies created coherent connections between the content of those reports, GPRA documents, and PART reviews.

Nor did many agencies try to show how their GPRA goals and PART targets connected. Also, the online posted versions of PART reviews were often disappointing because they lacked much of the information congressional staff and the public sought. For example, when the PART website cited an evaluation study, it often failed to provide a full citation or URL to make the study easy to obtain.

OMB has recognized the problem with the Performance and Accountability Report and has encouraged agencies to experiment with a short performance highlights document. But changing the document size will not fully address congressional and public frustration. Performance reports must deliver the information key audiences need.

GPRA, PART, and relevant OMB circulars do not ask agencies about target audiences for their performance information. The websites www.expectmore.gov and www.results.gov did not organize information to serve different audiences, despite progress on that dimension by many other federal agency websites. For performance measurement to be useful, and not just filler for documents required by law and OMB, agencies need to think more explicitly and strategically about who the key audiences for performance information are and how to meet their needs. They need to learn how to present information so it is understandable and useful to each target audience and how to confirm its use and usefulness. They also need to learn where and when to distribute it so key information reaches users in time to inform their decisions and actions. In short, every agency and program should think more explicitly about who needs what information to make better choices and improve performance. They should think explicitly about when they need it, where they need it, and in what format.

Principle Four: Structure Accountability Mechanisms to Encourage and Inspire, Not Embarrass, Reprimand or Punish

The accountability mechanisms of the federal performance management system—the expectations set and incentives used—are askew and need adjustment. As discussed earlier, the mechanisms OMB uses to motivate agencies focus on the wrong objectives: target attainment, PART ratings, and President’s Management Agenda Scorecard green lights. Focusing on these objectives implicitly punishes programs when targets are not met even when a program applies smart strategies based on available evidence, collects relevant data, and exerts strong effort. Moreover, OMB fails to take advantage of some of the strongest motivators for government workers, including a sense of accomplishment, genuine positive feedback, and ongoing opportunities to discuss problems and brainstorm solutions with other knowledgeable individuals.

GPRA sensibly requires agencies to set goals and measure performance toward them. GPRA does not, however, call for penalties for agencies that do not meet their targets.

PART introduced penalties. It penalizes programs with a lower total PART score when they fail to attain short-term ambitious targets and fail to make adequate progress toward long-term ambitious targets. Lower PART scores translate to a lower PART rating, which is treated as a proxy for agency performance in numerous venues. PART successfully increases agency attention to setting outcome-focused goals and measuring performance, but also creates a perverse incentive: it tempts agencies to pick timid targets they know they can meet, not ambitious targets more likely to stimulate the kind of innovation and energy that achieves higher performance levels.

The notion that it is problematic to penalize agencies that fail to meet their targets is somewhat counterintuitive. Nonetheless, it can be a serious problem. Misaligned accountability expectations can quickly compromise a healthy performance-improving dynamic. It is important for agencies and

programs to set specific targets, and it is stimulating when they select a few in priority areas that are ambitious. It is also important for managers to push their organizations to meet targets and to achieve continual performance improvement.

Yet if those targets are, in fact, ambitious, it is neither fair nor motivating to penalize people or organizations for not meeting them. It is discouraging and irritating. Penalties are useful for calling attention to problems that would otherwise be ignored and are necessary to assure attainment of standards (such as non-discrimination or caps on allowable releases of specific substances to the environment). Penalties levied for non-attainment of, or slow progress toward, an ambitious target are unfair, however, when intelligence, evidence, and effort have been applied. It is far better to hold agencies accountable for collecting and carefully analyzing evidence pertaining to problems, growth opportunities, and past experience, and for adopting and implementing cogent strategies to meet targets and improve performance.

Ambitious targets can be motivating, but only when they are also realistic given available skills, resources, and authority. Penalizing programs not meeting or making progress toward their targets when the programs request but do not receive changes in legislative authority can undermine the initiative of those otherwise doing their best. The PART process unfairly penalizes programs that lack needed legal authority to make program changes, even when a program proposes legislative changes to OMB and Congress.

The current performance system also makes insufficient use of several proven motivators: a sense of accomplishment, genuine positive feedback, and an opportunity to discuss problems with people who have relevant expertise to understand the evidence better and to brainstorm smarter strategies.

PART reviews assess and rate. They reward well-run programs with a good PART score and rating, but few OMB reviewers provide much constructive verbal feedback. Indeed, it might seem inappropriate for an OMB reviewer to praise a program manager,

since few OMB reviewers have enough program management expertise to provide program management guidance. The problem is that no one else provides that verbal feedback, either. Some OMB reviewers sensibly broker expertise when they see it is needed, but most assess, score, rate, and negotiate with agencies. Agencies are commended at meetings of the President's Management Council for "getting to green" on the President's Management Agenda Scorecard, but not recognized for improving performance trends.

The new administration needs to adjust the motivational mechanisms of the federal performance management system to reduce fear and the perceived unfairness in the system. It needs to adjust the accountability expectations and increase use of positive incentives that tap into intrinsic inclinations to do well and altruistic instincts to do good.