

RATING THE PERFORMANCE OF FEDERAL PROGRAMS

The measure of compassion is more than good intentions, it is good results. Sympathy is not enough.

President George W. Bush
April 2002

Federal programs should receive taxpayer dollars only when they prove they achieve results. The federal government spends over \$2 trillion a year on approximately 1,000 federal programs. In most cases, we do not know what we are getting for our money. This is simply unacceptable.

Good government—a government responsible to the people whose dollars it takes to fund its operations—must have as its core purpose the achievement of results. No program, however worthy its goal and high-minded its name, is entitled to continue perpetually unless it can demonstrate it is actually effective in solving problems.

In a results-oriented government, the burden of proof rests on each federal program and its advocates to prove that the program is getting results. The burden does not rest with the taxpayer or the reformers who believe the money could be better spent elsewhere.

There can be no proven results without accountability. A program whose managers fail year after year to put in place measures to test its performance ultimately fails the test just as surely as the program that is demonstrably falling short of success.

These ideas are simple, but they do not reflect how things have operated in the federal government. Instead, the Washington mentality has generally assumed that program funding should steadily increase with the only question being “by how much?” This type of thinking has wasted untold billions of dollars and prevented the emergence of results-oriented government.

This budget makes an unprecedented effort to assess the effectiveness of specific federal programs. It introduces a new rating tool to hold agencies accountable for accomplishing results. Programs are rated from Effective to Ineffective, and the ratings and specific findings produced will be used to make decisions regarding budgets and policy. The tool assumes that a program that cannot demonstrate positive results is no more entitled to funding, let alone an increase, than a program that is clearly failing.

This Program Assessment Rating Tool (PART) is discussed in greater detail below. It is far from perfect. In fact, after using it for the first time, we have identified a number of shortcomings that will need to be addressed. But it is an important next step in changing the way federal managers think about their responsibilities. It places the burden of proving effectiveness squarely on their shoulders. With further improvement and use, it will provide incentives for federal managers to make their programs more effective. It will provide meaningful evidence to the Congress and other decision-makers to help inform funding decisions, and identify flaws in underlying statutes that undermine effectiveness.

Demanding that programs prove results in order to earn financial support, however obvious and sensible, marks a dramatic departure from past practice. No one has asked about the extent to which Elderly Housing Grants help the one million very low-income elderly households with severe housing needs. Is \$4.8 billion in federal foster care funding preventing the maltreatment and abuse of children by providing stable temporary homes? Have federal efforts to reduce air pollution been successful? These programs seek to accomplish important goals but fail to provide evidence that they are successful in serving the citizens for whom they are intended.

Even programs known to be failing continue to get support. For example, the Safe and Drug Free Schools Program, which a 2001 RAND study determined to be fundamentally flawed, has only grown larger and more expensive. The current system discourages accountability, with no participant incentives to take responsibility, much less risks, to produce improvements in results.

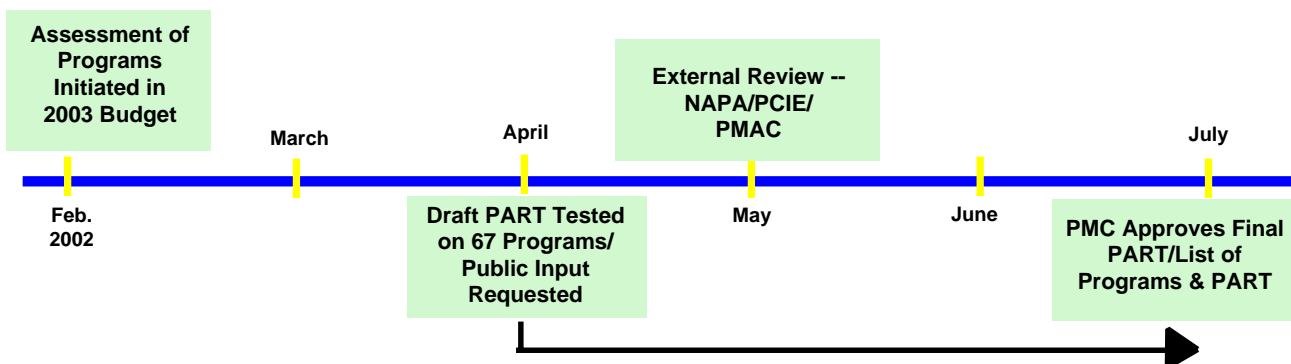
Previous administrations have grappled with this problem.

- President Johnson launched his Planning, Programming, and Budgeting System in 1966 to “substantially improve our ability to decide among competing proposals for funds and to evaluate actual performance.” The system was the first serious effort to link budgets to getting results, and a form of it remains in use at the Pentagon today.
- President Nixon followed with an effort called Management By Objective. This attempted to identify the goals of federal programs so that it was easier to determine what results were expected of each program and where programs were redundant or ineffective. Nixon stated, “By abandoning programs that have failed, we do not close our eyes to problems that exist; we shift resources to more productive use.”
- President Carter attempted to introduce a concept known as zero-based budgeting in 1977, to force each government program to prove its value each year. “[I]t’s not enough to have created a lot of government programs. Now we must make the good programs more effective and improve or weed out those which are wasteful or unnecessary,” he told the Congress and the American people in his 1979 State of the Union Address.
- President Clinton’s Administration also offered a broad agenda to “reinvent” government to make it cost less and do more.

Some of these efforts brought temporary improvement, and some vestiges remain. But the inertia of the status quo eventually limited the impact of each initiative, and we are no closer to measurable accountability than in President Johnson’s day.

Thus far, the most significant advance in bringing accountability to government programs was the Government Performance and Results Act of 1993. This law requires federal agencies to identify both long-term and annual goals, collect performance data, and justify budget requests based on

A Timeline of ...



NAPA = National Academy of Public Administration; PCIE = President's Council on Integrity and Efficiency.

these data. For the first time, each federal program was required to explicitly identify measures and goals for judging its performance and to collect information on an annual basis in order to determine if it was meeting those goals. Senator William Roth, sponsor of the measure, stated at the time that the act would bring about “accountability by federal agencies for the results they achieve when they spend tax dollars.”

Unfortunately, the implementation of this law has fallen far short of its authors’ hopes. Agency plans are plagued by performance measures that are meaningless, vague, too numerous, and often compiled by people who have no direct connection with budget decisions. Today, agencies produce over 13,000 pages of performance plans every year that are largely ignored in the budget process.

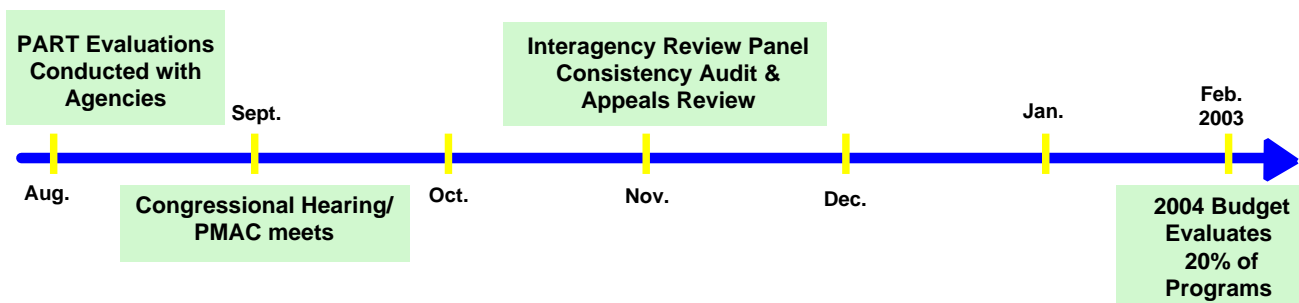
President Bush intends to give true effect to the spirit as well as the letter of this law. In the 2003 Budget, the Administration rated approximately 130 federal programs on their effectiveness. This first-ever attempt to directly rate program effectiveness was only a start and far from perfect. The criteria used to rate programs were not uniform and ratings were based on limited information. Its influence on budget decisions was limited.

This budget presents a new rating tool that, while still a work in progress, builds on last year’s efforts by generating information that is more objective, credible, and useful. Where a program has been rated, the results of the rating have often influenced the budget request. The accompanying timeline provides information on the development of the PART.

The PART is a questionnaire. The development of the PART began almost a year ago. The goal was to devise a tool that was objective and easy to understand. Most important, the findings had to be credible, useful, and ideologically neutral. For instance, the first draft PART questioned whether a particular program served an appropriate federal role. Because many believed that the answer to this question would vary depending on the reviewer’s philosophical outlook, it was removed so that the PART would yield a more objective rating. The accompanying timeline provides information on the development of the PART.

Whether a federal role is appropriate remains a critical consideration when determining a program’s funding level, and thus the issue is still central to budget deliberations. For example, the Manufacturing Extension Partnership program, which supports a nationwide network of centers providing technical assistance, earned a Moderately Effective rating, even though it is not evident that the federal government should be engaged in business consulting. Therefore, the budget continues the 2003 Budget policy of phasing out federal funding of mature centers with the goal of making them self-sufficient.

... the PART's Development



PMAC = Performance Measurement Advisory Council ; PMC= President’s Management Council.

How the PART Works

The PART evaluation proceeds through four critical areas of assessment—purpose and design, strategic planning, management, and results and accountability.

The first set of questions gauges whether the programs' design and purpose are clear and defensible. The second section involves strategic planning, and weighs whether the agency sets valid annual and long-term goals for programs. The third section rates agency management of programs, including financial oversight and program improvement efforts. The fourth set of questions focuses on results that programs can report with accuracy and consistency.

The answers to questions in each of the four sections result in a numeric score for each section from 0 to 100 (100 being the best). These scores are then combined to achieve an overall qualitative rating that ranges from Effective, to Moderately Effective, to Adequate, to Ineffective. Programs that do not have acceptable performance measures or have not yet collected performance data generally receive a rating of Results Not Demonstrated.

While single, weighted scores can be calculated, the value of reporting, say, an overall 46 out of 100 can be misleading. Reporting a single numerical rating could suggest false precision, or draw attention away from the very areas most in need of improvement. In fact, the PART is best seen as a complement to traditional management techniques, and can be used to stimulate a constructive dialogue between program managers, budget analysts, and policy officials. The PART serves its purpose if it produces an honest starting point for spending decisions that take results seriously.

An initial draft of the questionnaire was released for public comment in May 2002. It was reviewed by a number of independent groups, including the Performance Measurement Advisory Council, chaired by former Deputy Secretary of Transportation Mortimer Downey, and a group from the President's Council on Integrity and Efficiency. The PART was then the subject of a special workshop sponsored by National Academy of Public Administration and a congressional hearing. The President's Management Council approved a final version of the PART on July 10th and it was released for use on July 16th.

One lesson learned from these reviews, and past efforts, was not to try to do too much at one time. Thus, the Administration plans to review approximately one-fifth of all federal programs every year, so that by the 2008 budget submission every program will have been evaluated using this tool. For this year, OMB completed reviews for 234 diverse programs as a representative sample of government programs, as well as to test the flexibility of the PART. Chosen programs vary by type (such as regulatory, grants, or direct assistance), as well as size. To test how well the PART would confirm expectations, some programs generally

considered effective (such as the National Weather Service) were included, as well as some widely criticized as less effective, (such as compliance with the Earned Income Tax Credit (EITC)). Finally, several items of great interest to the President or the Congress were selected, such as programs scheduled for reauthorization this year.

The PART's approximately 30 questions (the number varies depending on the type of program being evaluated) ask for information which responsible federal managers should be able to provide. For instance:

- Is the program designed to have a significant impact in addressing the intended interest, problem, or need?
- Are federal managers and program partners (grantees, sub-grantees, contractors, etc.) held accountable for cost, schedule, and performance results?
- Has the program taken meaningful steps to address its management deficiencies?
- Does the program have a limited number of specific, ambitious long-term performance goals that focus on outcomes and meaningfully reflect the purpose of the program?

- Does the program (including program partners) achieve its annual performance goals?

The burden is on the program to demonstrate performance. Absent solid evidence to support a positive answer, the answer is deemed not favorable and the program receives a lower rating. The requirement of hard evidence fulfills the principle that federal managers must be held accountable for proving their programs are well designed and well managed.

Programs must also prove they are working. Programs that cannot demonstrate they are achieving results either because they have failed to establish adequate performance measures, or have no supportive performance data show a rating of Results Not Demonstrated. The other ratings are, in descending order: Effective, Moderately Effective, Adequate, and Ineffective. For more information on how the PART is used see the sidebar entitled “How the PART Works” in this chapter. Even greater detail is provided in the *Performance and Management Assessments* volume.

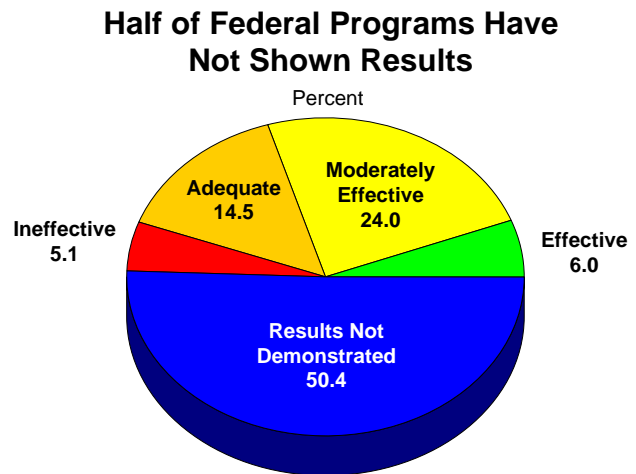
The purpose of the PART is to enrich budget analysis, not replace it. The relationship between an overall PART rating and the budget is not a rigid calculation. Lower ratings do not automatically translate into less funding for a program just as higher ratings do not automatically translate into higher funding for a program.

For example, the PART assessment found that the Department of Defense does not manage its Communications Infrastructure program consistent with industry’s best practices. The budget proposes to develop measures for department-wide evaluation and provides additional funding for this activity because of its importance to national defense. Budgets must still be drawn up to account for changing economic conditions, security needs, and policy priorities.

This first performance assessment process confirmed many longstanding suspicions. While the PART gives new programs a full chance to prove their worth, most rated this year have been in existence for years. Federal programs have inadequate measures to judge their performance. Over half of the programs analyzed received a rating of Results Not Demonstrated because of the lack of performance measures and/or performance data. The vast majority of programs have measures that emphasize inputs (such as the number of brochures printed) rather than outcomes or results. For instance, the Ryan White program previously only measured the number of people it served; in the future it will also measure health outcomes, such as the number of deaths from HIV/AIDS.

Patterns emerge in several broad areas of government activity. Despite enormous federal investments over the years, virtually none of the programs intended to reduce drug abuse are able to demonstrate results. Such a finding could reflect true failure or simply the difficulty of measurement, so further analysis is in order.

Overall, grant programs received lower than average ratings, suggesting a need for greater emphasis on grantee accountability in achieving overall program goals. Many other examples can be found in the PART summaries. The programs found to have inadequate measures will now focus on developing adequate measures and collecting the necessary data before the evaluations are done for next year. Just as importantly, programs that have not yet been evaluated can anticipate such scrutiny and assess the measures they currently have, and improve them where necessary.



The PART instrument, and the entire endeavor of budgeting for results, is obviously a work in progress. We strongly encourage individuals or organizations to examine the worksheets of programs of interest to them and provide comments on how the evaluation of that program could be improved. Is there evidence that was missed or misinterpreted? Do the answers seem consistent with the individual's or organization's experience?

Despite an extensive review process and refinements, the tool itself still has limitations and shortcomings. This is not surprising in its debut year. But these issues will need to be addressed before the process begins for next year. We welcome comments and proposals addressing all these issues. Comments can be emailed to performance@omb.eop.gov or mailed to the Office of Management and Budget/PART Comments, Eisenhower Executive Office Building, Room 350, Washington, D.C. 20503. These issues include:

- *Increasing Consistency.* A sample of PARTs was audited for consistency by an interagency review panel, and some corrective action was taken based on its findings. Nonetheless, similar answers were too often subject to different interpretations. For instance, in the majority of cases, if the current goals for performance were found inadequate, the program was given little or no credit for meeting them. How can consistency of answers be improved?
- *Defining "Adequate" Performance Measures.* Developing good performance measures is very difficult. The criteria for good measures often compete with each other. For instance, it is preferable to have outcome measures, but such measures are often not very practical to collect or use on an annual basis. The fact is there are no "right" measures for some programs. Developing good measures is critical for making sure the program is getting results and making an impact. Given the widespread need, we are already planning to provide federal officials more training on how to select good performance measures. We ask, what else should be done to improve the development of consistently better performance measures?
- *Minimizing Subjectivity.* While intended to be based on evidence, the interpretation of some answers to the questions will always require the judgment of individuals and therefore be somewhat subjective. How can the process be modified to assume greater balance and objectivity?
- *Measuring Progress Towards Results.* Since the PART focuses on results, the rating provides little credit for progress already being made – improvement that may not be reflected in program results for a year or more. For instance, for the National Park Service to improve the condition of its facilities, it must first inventory them and prepare a plan for eliminating the maintenance backlog. Those necessary initial steps do not represent results, as measured with outcomes, and get little credit in the PART. How can the tool or its presentation better capture progress toward results?
- *Institutionalizing Program Ratings.* If they are to drive continual improvement, program ratings must be conducted regularly so that credit is given when earned and deficiencies are identified in a timely manner. Toward that goal, the ratings were built into the budget process. How can this activity be continued without requiring an elaborate infrastructure for it to be sustained?
- *Assessing Overall Context.* As the General Accounting Office pointed out, the PART does not capture whether a program complements other, related programs or whether it may work against the goals of the Administration's other initiatives. How can the PART rating better consider programs in a broader context?
- *Increasing the Use of Rating Information.* The PART may or may not prove to influence budget decisions made by the Congress. But it certainly will give lawmakers more detailed information on which to base their choices. Most people understand that wise spending decisions go beyond the widely discussed increases or cuts in one program or another. For instance, a poorly run

program can benefit from a budgetary boost to address its identified problems. Or a popular program can flourish without an increase, because it is steadily increasing its productivity and, therefore, its service results.

Taken seriously, rigorous performance assessment will boost the quality of federal programs, and taxpayers will see more of the results they were promised. What works is what matters, and achievement should determine which programs survive, and which do not. The public must finally be able to hold managers and policymakers accountable for results, and improve the performance of federal programs in the way that Presidents and Congresses have sought for decades. Whether resources shift over time from deficient to good programs will be the test the PART itself must pass.