

Playing the Wrong PART: The Program Assessment Rating Tool and the Functions of the President's Budget

Joseph White is the Luxenberg Family Professor of Public Policy in the Department of Political Science at Case Western Reserve University. His research focuses on federal budget policy and politics; health care policy and politics in the United States and other advanced industrial democracies; policy consequences of an aging society; and policies to control health care costs. He is the coauthor with Aaron Wildavsky of *The Deficit and the Public Interest: The Search for Responsible Budgeting in the 1980s*.

E-mail: joseph.white@case.edu

The extensive literature about the George W. Bush administration's Program Assessment Rating Tool (PART) has not emphasized an issue that appears quite clearly in interviews with senior Office of Management and Budget (OMB) and congressional staff. Budget analysis can pursue efficiency in two ways. One, the object of traditional analysis, involves assessing programs to understand the ratio of inputs to outputs within an agency, and thus the effects of more or less funding. Another approach, termed "budgeting for results," measures program performance so as to allocate funding among programs in a way that increases total welfare. The second approach is much more difficult because it necessitates comparison of measures of unlike phenomena, requires expertise that often does not exist, and is more easily contested as invoking values rather than facts. Both congressional and OMB sources report concerns that PART weakened budget analysis by diverting resources from traditional analysis. If one goal of reformers is for the OMB to provide analysis that will influence Congress, the focus should be on strengthening the capacity to do traditional budget analysis.

What is the role of the president's budget within federal budgeting? In particular, how does the formal presidential budget process, as managed by the Office of Management and Budget (OMB), contribute to the success or failure of the federal budget process?

In the summer of 2008, I decided to revisit this question, which I had studied nearly two decades earlier (White 1991). In addition to reviewing the literature produced over that period, I conducted new interviews with 22 mostly senior officials and former officials who had clearly relevant and usually quite extensive experience in the budget process. Among them were two people who had served as director of the OMB; two directors of budget review (the senior career position in the OMB); five who had served as division associate directors (one of the eight civil servants who report directly to the politically appointed program associate directors); and two

who had served as clerk (staff director) of the House and Senate appropriations committees. Of the seven congressional sources, five had worked at the OMB.¹ These interviews were part of the basis for an overview article (White 2009). But they also provided perspective on a question that has provoked substantial controversy in recent public administration and budgeting literature: the performance of the George W. Bush administration's Program Assessment Rating Tool (PART).

PART's advocates described it as a reasonable effort to improve the efficiency of the federal government (see, e.g., Blanchard 2008; Breul 2007; Redburn et al. 2008, Shea 2008). Other analyses provided more mixed reviews but see some positive aspects in its design (e.g., Dull 2006; Frederickson and Frederickson 2006; Gilmour 2007, 2008; McMurtry 2005; Metzenbaum 2009; Nathan 2005; Posner and Fantone 2008, Wichowsky and Moynihan 2008). Still other accounts were much more negative (e.g., Gueorgieva et al. 2009; Nathan 2008; Radin 2006, 2008), but, on balance, PART appears to appeal to a deeply entrenched desire within the public administration community to find a way to budget "by performance" or "for results."

I did not set out to study PART, but it was emphasized prominently by the OMB respondents when I began asking about how the presidential process had changed. "During the Bush administration," one told me, "the president's management agenda and interacting with that agenda has been a major change. Implementation of PART has been a major role." "PART we spend an enormous amount of time on," another source explained. Therefore, I asked about PART in subsequent interviews.

Congressional respondents were mostly negative, although they said that the assessments occasionally provided some useful information. This should not surprise readers who are familiar with the literature on what Richard Nathan called "the alphabet soup

of systems like PPBS, MBO, ZBB, NPR, and GPRA" (2005, 207; for the first three, see Schick 1966, 1973; Wildavsky 1966, 1969, 1981). As described later, their responses were in line with those reported in other sources (GAO 2008; Metzenbaum 2009). The credibility of the criticisms, however, is significantly enhanced by the fact that the overall view from senior OMB career staff was neutral at best: some positives were identified, but there was extensive doubt as to whether the contributions from PART were worth the time spent on it.

There are important reasons for understanding both the role of the president's budget within federal budgeting and the ways in which the budget process may increase the efficiency of federal operations. Presidents and their staffs may see the presidential budget only as a way to serve the president's interests; but it would not have been created and would have little influence separate from the president's powers of the veto and as a party leader if it did not serve some useful function for Congress as well (Caiden 1987; Mosher 1984; Rubin 1994; White 1991). The OMB can contribute in a way that is seen as useful even across party and institutional lines by providing analysis that identifies how well agencies can use increments or the consequences of decrements. Information about the effects of proposed changes to agency budgets can help decision makers pursue efficiency, defined as a better ratio of outputs to inputs for a given agency. "Scrubbing the estimates" at the agency level is the core function of traditional budget analysis.

As an approach to budgeting, PART in particular and "budgeting by results" in general emphasize comparisons across units rather than allocations within them. The assumption is that programs with better results should—multiple caveats aside—be favored; a portfolio of better-performing programs will result in a more efficient government. Yet comparison across programs is much more problematic than evaluation within programs because programs serve different goals. How PART actually was implemented gave appropriators further reason to distrust the analyses. At the same time, they complained that they were not getting from the OMB the traditional analysis that they wanted and needed. The OMB sources agreed that the work needed for PART diverted resources from traditional analysis.

The next part of this essay describes the concerns about the usefulness of PART. The following section explains how those concerns fit with an historical perspective on the purposes of the executive budget and how budget analysis can be used to increase efficiency.

PART, the OMB, and the Appropriators *The PART Design*

PART was the Bush administration's attempt to improve the process created by the Government Performance and Results Act (GPRA) under the Bill Clinton administration. The GPRA was one of a series of initiatives during the 1990s whose sponsors hoped to improve federal management, thereby making the government more efficient and reducing the federal deficit (Dull 2006; GAO

1997; Kasdin 2010; McMurtry 2005; Posner 2007; Schick 2004; Thompson 2002; Tomkin 1998). The law required that agencies create performance plans and that the OMB oversee this process. The plans, in theory, would create standards that both would give incentives for better management and would provide a basis for assessing performance.

In practice, the GPRA had many flaws. Legislators had conflicting views of its purposes. While some legislators may have liked the idea that agencies would clarify their programs' objectives so as to assess performance as a general principle, defining objectives is a task that congressmen tend to believe is their job if they care about a program, and so easily can get agencies into trouble. Agencies also were suspected of choosing measures to make themselves look good. At the same time, the GPRA could be criticized for not requiring anyone to pay attention to the reports (Dull 2006; Kasdin 2010).

As a result, while in some agencies, the effort seemed to some participants to be worth the trouble, and in Congress, observers occasionally found some of the data useful (Metzenbaum 2009), the GPRA did not come close to meeting its supporters' expectations. In the classic pattern of such reforms, many appropriators found the whole exercise, as one senior aide expressed it to me, "ludicrous" and made hardly any use of it. During the Clinton administration, the fact that the reports were being performed by an administration of the opposite party did not help, but the fact that the agencies got to define their own standards also seemed problematic to the appropriators. Another veteran appropriations aide referred to the GPRA as "the biggest waste of paper in Washington."

The assumption is that programs with better results should—multiple caveats aside—be favored; a portfolio of better-performing programs will result in a more efficient government. Yet comparison across programs is much more problematic than evaluation within programs because programs serve different goals.

With George W. Bush's election, both the "public management community, committed to the ideas of performance management and budgeting, and the conservative think tanks" wanted the GPRA process strengthened (Dull 2006, 201). To solve the perceived problems that the GPRA reports were not comparable and required no action, the OMB created the Program Assessment Rating Tool. As an executive mandate, PART could be enforced through executive processes. It required strategic planning but put greater emphasis on measures of "performance," variously defined. But perhaps the most important practical difference between the GPRA and PART was that the OMB examiners now made final determinations about both the questions and the answers.

After a widely criticized first effort for the fiscal year 2003 budget, the OMB in 2002 designed a new instrument, with feedback from outsiders such as the Government Accountability Office (GAO) and the National Academy of Public Administration, that first was implemented during the budget review in 2002 for the fiscal year 2004 budget submission. About 20 percent of programs were "PARTed" each year, so by 2007 (the fiscal year 2008 budget), 976 programs representing 96 percent of the budget had been assessed (Redburn et al. 2008).

Under PART, the OMB examiners were to rate programs as “effective,” “moderately effective,” “adequate,” “ineffective,” or “results not demonstrated.” Scores consisted 20 percent of an assessment of whether the program design was logical and purpose clear; 10 percent of whether the agency had a strategic plan with valid goals; 20 percent of ratings of various crosscutting aspects of program management, such as financial oversight; and 50 percent of ratings of performance based on measures of stated goal achievement. This design could be criticized for evaluating programs in part on standards that were not performance at all. Critics argued that some of the measures, such as clarity of goals, were biased against certain types of programs, such as block grants, for which part of the point is to leave goals largely to the discretion of the grantees (Nathan 2008; Radin 2008; Stalebrink 2009). Even the idea that agencies should be held responsible for performance may be questioned for programs in which the federal government is intended not to have operating responsibility (such as grants), so the accountability chain is designed to be weak (Nathan 2008).

Logically, “effective” had to mean something similar for all programs; the basic categories reflected a view that certain questions are relevant to assessing any program. At the same time, programs differ, and supporters of PART argued that it recognized this fact because the questions differed somewhat across seven categories of programs (Blanchard 2008). The categories included block/formula grants, capital asset and service acquisition, competitive grants, credit programs, direct programs, regulatory-based programs, and research and development.² Starting with a set of 25 questions that would be asked of direct programs, the instrument provided further questions that could be asked of different kinds of programs. However, unless the examiner chose to say that some of the standard 25 questions were “not applicable” to a program, about 95 percent of the score for a block grant, competitive grant, or credit program would be based on questions also asked of direct programs. And 88 percent of the score for a capital program, 83 percent of the score for a regulatory program, and 77 percent of the score for research and development programs would be based on the same measures used for direct programs.³

Hence, the instrument’s “flexibility” depended greatly on the decisions made by individual examiners. The examiners also, ultimately, determined the measures used to answer the questions. Thus, in the view of one respondent within the appropriations process, “the problem early on was they basically let individual OMB examiners devise it any way they wanted.”

Other research has reported the same pattern from the agency perspective. As a report by Shelley Metzbaum put it, “Agency staff identified reviewer inconsistency as a serious problem” (2009, 35). One of her respondents, for example, reported that, “I have worked with five different examiners, 14 programs, 19 program reviews with re-PARTs, and 60 measures, plus 40 or 50 follow-up actions. It is amazing to me how they take the guidance and look at it differently ... It is mostly dependent on the examiner. That is a serious problem with the PART” (2009, 32). John Gilmour provided another indicator of

this discretion in his discussion of the extent to which examiners insisted on having true outcome measures. He cited examples, ranging from the Department of Energy’s program to ensure safe transport of nuclear materials to the State Department’s consular services and the Bureau of Labor Statistics, for which examiners accepted output measures instead of outcome measures. But Environmental Protection Agency (EPA) programs “were held across the board to end outcome measures,” even though that is especially questionable for those programs (Gilmour 2007, 28). As one of my OMB respondents explained, “the problem with EPA is, the more you go to outcome measures, the less control they have over it. Because they don’t implement the programs on the ground. We pretty much know what we want to do on air pollution, and EPA will write the regs. But there’s a huge lag ... you don’t see the impact for most of a decade. You can conceptualize the measures, but it’s harder to track progress.”

Supporters of PART argued that the examiners’ discretion was a good thing because the evaluations “represent the combined wisdom of career officials. This is not a political process” (Senator Wayne Allard, quoted in Blanchard 2008, 71). This definitely was not how it looked from the agencies. A GAO report explained,

[C]oncerns exist among federal managers regarding the quality of OMB’s assessments. Specifically, managers responding to our survey expressed concern that OMB examiners may be spread too thinly and do not have sufficient knowledge of the programs ... the suggested improvement to PART with the highest level of endorsement from federal managers familiar with PART was to ensure that OMB’s examiners have an in-depth knowledge of the programs they review. Seventy percent of respondents indicated that this was a high to very high priority for improving PART. (2008, 21)

Perhaps the most striking aspect of my interviews is that veteran OMB personnel reported similar concerns. In fact, they argued that PART was most useful as a way to educate *inexperienced* examiners about the programs they reviewed. As one put it, “for my junior examiners coming in, I think PART is very useful in terms of having a structured way to learn the program. It asks a lot of the right questions. It’s too structured for my taste, and for the senior examiners would not be useful, but we have a lot of examiners coming in from grad school and it’s good when you’re starting with a relatively low knowledge base.”

The examiners’ discretion was a key reason why appropriations committee aides, who often are far more experienced than the examiners, did not trust the assessments. As a departmental budget officer saw the situation, “it’s something that was layered on top of the responsibilities of the examiners. Their ability to do this in a particularly useful way is limited, and as a result what’s been produced has been a mixed bag. And my sense is that the utility of it on the Hill in particular is reduced not only because the quality is various but because the secretiveness of the process, the black box with things popping out, does not engender a lot of confidence among outsiders in the result.”

The examiners’ discretion was a key reason why appropriations committee aides, who often are far more experienced than the examiners, did not trust the assessments.

PART advocates believed that having the standards published made the process “transparent,” but a negotiation between an OMB examiner and agency officials in which the examiner has the final decision is not, to outsiders, an open process. An OMB respondent who has written in favor of PART emphasized that “you or anyone else can go on the website and see what’s what, at least as seen by OMB examiners.” But there is little reason for veteran staff in Congress, or legislators, to credit such broad judgments when made by individual examiners. The GAO reported that the absence of consultation between OMB and congressional staff to either determine the evidence used for analysis or set priorities as to which programs to analyze was one reason for “most congressional staff we spoke with not using the PART information” (2008, 19).

Evidence about Benefits from PART

PART contained two aspirations for performance budgeting. One was that either the inability to document performance or measures that show poor performance could lead agency management to commit to change. PART did give agency staff strong incentives to agree with their examiners about a set of measures. They did not, as one OMB respondent recorded, want to be downgraded on “their management scorecard” within the President’s Management Agenda for having no rating. But an incentive to agree on measures is not the same as an incentive to improve performance.

The GAO’s surveys of federal managers on performance and management issues show that PART likely contributed to an increase in the number of performance measures available within the federal government. The same surveys, however, show no change or even a decline between 1997 and 2007 in the extent to which managers reported making substantial use of those measures for management activities (GAO 2008, 4, 6). It seems fair to say that PART did not create a culture of managing for performance. What it did do was give managers at the upper levels of agencies strong incentives to make their programs look good on the measures.

When Gilmour studied agencies that raised their ratings, he found that they “explained their success with almost identical language. ‘We learned how to take the test,’ they said.” They learned “how to write more acceptable answers” and “how to devise acceptable outcome measures.” None of his respondents “claimed they had introduced significant management improvements or changed program design to raise their rating” (2008, 29–30).⁴ The evidence does not support the hope that measurement of results would produce organizational transformation (for further analysis of the difficulties, see Kasdin 2010).

The second goal was that measurement of results could “provide decision-makers with information needed to better allocate scarce resources in a way that will yield the greatest benefit” (Gilmour 2008, 23). Did this mean cutting programs that received low ratings? Advocates of PART sent mixed messages on this point (Breul 2007; Dull 2006). Nevertheless, PART clearly was designed from a premise that programs needed to be challenged, and it was interpreted by the administration’s allies as a way to cut spending.

It seems fair to say that PART did not create a culture of managing for performance. What it did do was give managers at the upper levels of agencies strong incentives to make their programs look good on the measures.

The OMB guidance stated that the process “requires a high standard of evidence and it will likely be more difficult to justify a *Yes* than a *No*” to the question of whether a program is effective (Dull 2006, 203). Lloyd A. Blanchard, who as program associate director for the General Government Division of the OMB in 2001 was an early advocate for the PART process, argued that public organizations seek to “enhance” their budgets. Because “there is no natural incentive to promote efficiency and effectiveness in government program service delivery,” through PART, “an incentive can be created by rewarding better-performing programs with larger increases in their budgets, and punishing poorer performers with smaller increases or decreases in budgetary allocations” (Blanchard 2008, 72). In short, “budget maximizing bureaucrats” must be opposed or co-opted.⁵ Appropriations staffers who believed that the purpose of the process within the Bush administration was mainly to create rationales for cuts had good reason for their perception.

Did PART affect allocations among programs? The most positive respondent reported that the PART analyses “were used for presenting alternatives to the leadership ... that doesn’t mean decisions necessarily were due to the PARTs, but it was used and became part of the vocabulary certainly.” That begs the question, however, of whether PART scores were driving analysis or simply reflected preexisting bias within the OMB, which would have caused the same decisions within the presidential process. One has to be rather skeptical, after all, when one sees that “programs established under Democratic presidents received systematically lower PART scores than programs initiated under Republican presidents” (Gilmour and Lewis 2006, 743), or that redistributive programs received significantly more negative scores in a Republican administration (Greitens and Joaquin 2010).

Gilmour and Lewis attempted to separate out any bias effects with an instrumental variables technique and concluded that “PART scores were positively associated” with OMB proposals for “Democratic programs but not the rest” (2006, 742). They concluded that “it appears to be easier to implement performance budgeting with programs that one does not support” (751). Blanchard, using more years of data, found a larger effect on allocations while simply dismissing the concern about underlying bias. His evidence about effects on actual budgets is not strong. The pattern in the early years is inconsistent (before 2007, the highest-performance quartile did worse than the next highest), and the percentages of variance explained were very small: 4 percent or less with no controls, and 7.1 percent with a variety of other factors such as the size of the program, program type, and whether the department was part of the “fight against terrorism” added (Blanchard 2008, 78–87).

Asked directly for examples of the PART exercise affecting OMB budget proposals, my respondents described it as at best a minor factor. A division associate director reported that a program that was not of much interest to a department “turned out to save money; OMB was able to use their PART score to justify an add.” Another program got a low score, though it was already under fire (“they might have tried to kill anyway, but the low PART

score helped"). The same director, however, reported that career staff "felt the way we handled it ... was not the right approach." Another senior official observed that the examples tended to be "at a micro level," such as some watershed programs—programs that were not getting a lot of attention in the first place. Asked directly, "was it worth all the work?" this person responded, "I will refrain from answering that question ... In essence the PART has some value, but you can question whether it ought to be applied to all programs."

Both the OMB and congressional respondents suggested that PART did not work well for the National Institutes of Health, for which it is particularly difficult to define achievable goals and for which "there are so many diseases" that comparison is unmanageable—in spite of the somewhat different standards for research programs. Some PART "findings" were not news at all. The president's budget for fiscal year 2004 reported that "[d]espite enormous federal investments over the years, virtually none of the programs intended to reduce drug abuse are able to demonstrate results" (OMB 2003, 51). Appropriations staff (both Republican and Democratic, in both the House and Senate) had been telling me the same thing in interviews since the late 1980s.

My interviews suggest that there were occasional cases of the analysis influencing presidential submissions, which, in turn, can influence appropriations decisions directly—especially when Congress and the presidency are controlled by the same party. Both the available studies and the interviews, however, suggest that any such influence was quite marginal. Perhaps a more important question, then, is whether the work done on PART gave, as an early OMB explanation of the process promised, "lawmakers more detailed information on which to base their choices" (OMB 2003, 52). In fact, legislative staff reported the opposite, and the observations from within the OMB supported that concern.

The Costs of PART

PART was viewed by appropriators as having both credibility problems and opportunity costs. Some of the credibility issues were attributable to distrust: "the sense was that the program evaluation process was established for the purpose of destroying programs, so there wasn't a great deal of respect for those evaluations." The fact that PART scores basically were determined by individual examiners did not add to their credibility. But veteran appropriations staff also emphasized the opportunity cost. One of my respondents commented that the PART analyses "have some good data but it's sort of crowding out other information. We're not getting much of a sense of how our agencies are doing."

Appropriators want to know what would be bought with more money or lost if the agency received less money. Their basic worldview was summarized neatly by a ranking House appropriations subcommittee member in the 1980s when I asked him how appropriating was different from authorizing. Appropriating, he said, was much simpler because there were "three questions: What is it? Why do you need it? And if it's so important, how did you do without it last year?" The trouble with PART for appropriators is that it does not answer those questions. As the House Appropriations Financial Services and General Government Subcommittee complained in the 2008 committee report on its bill,⁶

For years, the Committee has directed departments and agencies to improve the budget justification document quality and presentation by including relevant and specific budget information. While the Committee has seen some improvement in a few submissions, most justifications continue to be filled with references to the Program Assessment Rating Tool (PART), drowning in pleonasm, and yet still devoid of useful information ... The Committee finds little use for a budget justification which does not reveal specific details of the measurable indicators and standards used to evaluate a program's performance, relevance, or adherence to underlying authorization statute. Further, the Committee has little patience for agency heads who cannot explain the rationale behind a program's funding level other than "the PART score," "getting to green," or "this is what OMB provided." (U.S. House 2008)

Metzenbaum's respondents made much the same point. As one put it,

No one has a problem with the agency showing us the overall picture, but we cannot find information that tells us this is how they spent the money and how they are planning to spend the money in the coming year if appropriated. Without that core, we cannot go forward ... You need to respect the critical core of what the budget is, and how you would propose to spend it, which would be part of a conversation about how that program would contribute to a goal ... It is fine to present budgets with respect to performance, but not at the expense of core information that staff and the public need. (2009, 26–27)

The OMB veterans shared the concern about the opportunity cost of PART. The basic question may be phrased as, does PART improve budget analysis? The overall answer from my OMB respondents is that it could make a contribution, but not if it displaces traditional analysis, and not particularly if the examiners are knowledgeable to begin with. The earlier quote about PART being more useful for less experienced examiners expresses part of this perception. Another source called PART "a structured approach to doing something good examiners were doing all along." Yet the most common response about PART was that, in one senior career official's words, it "really ate a lot of time." This was viewed as a significant cost.

"One thing I used to do," one division associate director recalled, "was sit down with examiners and identify two or three things we would need to learn more about from March to September ... we could no longer do that, because PART sucked up every moment between February and September." In the late 1980s, the amount of time spent dealing with Congress was viewed as making it difficult for examiners to learn about the actual operations of agencies (White 1991). By 2008, the OMB was much less directly engaged with Congress (White 2009), but, in one career veteran's words, there had "been a gradual expansion of legislated tasks, and the result is that there is a substantial amount of examiners' time that gets consumed with responding to requirements that are either mandated or legislated on the institution. During the Bush administration the president's management agenda and interacting with that agenda has been a major change."

PART and other management initiatives, therefore, diminished the capacity to do traditional budget work. The problem was made worse by the erosion in OMB staffing. "Most of us who've been here a longer time ... believe that the quality of our thinking and analysis is probably less than it used to be. Not because we don't have a lot of really smart people here but because they're spread so thinly across so many activities, and the number of people in OMB has gone down." OMB veterans worried that examiners knew less and less about programs. As one put it, "they've cut the hell out of OMB's budget in the last couple of years. The budget constraints absorbed within OMB have made it very difficult to continue to have field education for examiners."

Another senior career official expressed this concern most strongly. He argued that there had been a "slow erosion" of capacity to go in the field and really learn about the programs that had "gotten more critical" in recent years. "When examiners can't tell if an agency is lying because they've never been out there to check things," he explained, "the whole relationship has fundamentally changed. They say whatever they need to say to OMB, because the examiners can't tell. To me that's been one of the problems with the President's Management Agenda. There's a whole lot of paper there, but we don't have the time to see if any of it is actually representative of anything. It's way too much."

The workload from PART added to the obstacles created by high turnover, budget constraints, and, therefore, a smaller, younger, less experienced staff. There have been other times when the ability to do the work was a concern, and OMB staff proudly insisted they would handle it (Berman 1979; Tomkin 1998). But the concern certainly seemed strong as of 2008, and senior staff less confident. In one respondent's words, "our ministerial functions we could always do. Now we're so low ... [a colleague] says he gets scared when an apportionment comes in because he's not sure the staff know how to do it right."

The Role of Executive Analysis in Congressional Budgeting

The concerns about PART from both appropriators and OMB should not be taken to suggest that the president's budget during the George W. Bush administration had little effect on congressional budget decisions. The president has power both as a party leader and as the wielder of the veto. Hierarchical norms enable the OMB to limit (though only partially) the flow of information between agencies and Congress that would contradict the preferences expressed in the president's budget. Whether it succeeded or failed, PART would not change these aspects of presidential influence on spending decisions.

At some times, the presidential budget apparatus has functioned essentially as staff to support the administration in a war with Congress to capture budgetary territory (Hecklo 1984; Johnson 1984; Matthiasen 1988; Tomkin 1998; White 2009). But when an executive budget process works well, it serves not only to advocate for the president's budget but also to help Congress do its own analysis.

Appropriators might be able to convince their colleagues in Congress that Agency X does not need an increase, or can live with a Z percent increase or decrease, if the appropriators can say that they have data about the operational needs of Agency X. Traditional analysis asks a micro-level performance question: in essence, what would be the result of giving more or fewer resources of certain types to an agency? (Tanaka, O'Neill, and Holen 2003). It is quite another matter to give two programs PART scores and say that one is more deserving than another because it earned a higher rating, especially when the appropriateness of the indicators is questionable and the scores are assigned by individual budget examiners whose authority of expertise (never mind position) is not recognized by outsiders.

The mismatch between executive analysis and congressional needs was evident to agency budget officers. What is an agency to do when the OMB insists that it emphasize performance information, including PART, and the appropriators want traditional justifications? (Kamensky 2007). The answer, as the director of the Budget Service for the Department of Education presented it, was to give

[E]ven my OMB respondents, on balance, believed that PART should be a supplement to traditional budget analysis, and that means that, if PART significantly reduces the time needed for traditional analysis, it is not a good investment.

the appropriators all the traditional material. Performance information could be included in narratives, as it normally has been when agencies had some, but with descriptions of the data sources. Appropriators preferred that "performance budget" information be submitted in a separate tab (Skelly 2006). Appropriators always will take more information, and they may use it if it is credible and helps them meet the political challenge of passing their bills. But even my OMB respondents, on balance, believed that PART should be a

supplement to traditional budget analysis, and that means that, if PART significantly reduces the time needed for traditional analysis, it is not a good investment.

The most critical reviews of PART have argued that it involves an approach to budgeting that ignores politics. PART advocates, Beryl Radin summarized,

are attracted to the dimensions of budgeting that flow from the internal management approach, including a reliance on the technical skills of the budget staff, a belief in a rational resource allocation model, a belief that information is neutral and can be determined to be true or false, a search for apolitical or at least bipartisan approaches, an emphasis on efficiency norms, a tendency to use the private or business sector as a model, and a strong reliance on the executive branch of government. (2008, 116–17)

These criticisms are all significant, but the contrast that Radin identifies should not be interpreted as meaning that the executive cares about efficiency norms and the legislature does not.

The rationale for PART treats the federal government as a portfolio of programs and reframes the budgeting question as, "in which of these programs would investment yield the most return?" As soon as analysis switches from how resources would be used within an individual program to comparisons among programs, however, it ceases to be value neutral. Because almost all programs serve

different values, efficiency comparisons across programs are, in practice, value choices. Not only are such comparisons easily criticized, but also they are harder to do credibly than evaluation of an agency work plan. Congress, naturally, is not interested in analysis that creates value judgments based on a framework that looks consistent but quite clearly, if one looks at the details, is based on individual budget examiners' extensive discretion.

The Budget Bureau's Traditional Contribution to Legislative Budgeting

I am not suggesting that individual members of Congress always are pleased to be told that their pet programs cannot use more money. There always have been tensions between the presidential budget organizations and the legislative budgeters—particularly the budget bureau and the appropriators—because the budget process is one battle in a contest for control of federal agencies. Yet *the appropriators need reasons to say no*. The executive budget process as it developed particularly over its first 50 years served this congressional interest in countering agency and agency advocates' claims about how much their programs "need."

This is not the place for a full review of the history of the federal government's executive budget (good monographs include Berman 1979; Mosher 1984; Tomkin 1998; for a congressional perspective on the 1921 act, see U.S. Senate 2008; for other sources, see White 2009). As Caiden (1987) argued, the core ideas of the executive budget movement fit poorly with the constitutional design, yet the procedures actually adopted better fit that design. The president received new and significant "first mover" powers, but Congress retained its ability to "dispose" of the presidential proposals.

The president's budget helped Congress by deflecting some blame. If the president proposed restraints that Congress was willing to accept, legislators could accept the proposal and blame him. Even if this only reduced their share of the blame, that was better than having the blame all to themselves. If Congress rejected a presidential proposal, and so had to replace it with another in order to keep spending within some target, legislators at least could get some credit from the interests they protected.

The second contribution made by the presidential budget process was more technical: it provided oversight of agency proposals and thus a way to "scrub" them to reduce "fat" and increase efficiency within programs. As Simon (1997) and Wildavsky (1966) both explained, "efficiency" is not the same as "economy," the goal of simply reducing spending. They defined efficiency as the ratio of outputs to inputs, so that it can be increased either by raising the output for a level of input or by lowering the input for a level of output. False economy in inputs could reduce outputs even more than inputs, making the organization less efficient. "Efficiency" in this sense is not a politically loaded concept, and so it can be useful to political actors who otherwise might be in conflict with each other.⁷

As Schick explained, "[e]xpenditures for running operations and services are ... the areas that budget makers look to for short-term savings to align revenues and expenditures" (2004, 96–97). In the United States, this has been true of both the budget bureau and the appropriations committees, so if the budget bureau can help

identify such efficiencies, that is often a service to Congress. Different politicians would put programmatic efficiencies to different uses, but this only shows that increasing programmatic efficiency is the closest thing to a neutral competence. Thus, in one of my interviews two decades ago, a senior career official said that "the idea I grew up with was to be a neutral competent budget analyst. As I told my staff, that meant *if it was a Republican administration trying to minimize cost; if it was a Democratic administration, how to maximize value for the money we had.*"

Even now, appropriators will say that having a presidential review is necessary. The best example of why is the National Cancer Institute's "bypass budget." The institute legally is allowed to submit its true preferences directly to Congress. When asked directly whether that is useful, an appropriations aide replied, "No. It's just too expensive, nobody can use it."

The budget bureau across its history sometimes has pursued another way to improve efficiency within agencies: improving agency management. The relationship between budgeting and management competence, however, is particularly controversial (Berman 1979; Dickinson and Rudalevige 2007; Mosher 1984; Posner 2007; Radin 2008; Stone 1990; Tanaka, O'Neill, and Holen 2003; Tomkin 1998). Putting PART within the context of the OMB's struggle to define a plausible management function would require another article. Because my focus was on budget formulation, it was not a topic in my interviews. There is little reason to believe, however, that PART did much to improve federal management. The best that can be said is that it increased the number of measures (GAO 2008; Gilmour 2008; Metzenbaum 2009).

The process of proposing and then appropriating budgets for what are now called "discretionary" programs is not, of course, the only part of federal budgeting that matters. It has become a smaller part, in both dollar and procedural terms, of budgeting over the past half century (Schick 1980; White 2009; White and Wildavsky 1991). Nevertheless, throughout the past decades of conflict and procedural change, a large part of the work of budgeting has focused on allocations of discretionary spending to individual agencies. It remains by far the largest part of the OMB's activity. Throughout a quarter century of interviewing in the appropriations committees, appropriators continually have expressed to me their interest in having a president's budget that provides information that they can use to figure out how to get the most value for money out of individual agencies. PART did not help.

Conclusion

So what, now, for PART in particular, and for "budgeting for results" in general?

PART now has been added to what Blondal, Kraan and Ruffner called the "long, if irregular history" of performance budgeting in the United States, a history that, as of 2003, had "failed to make a substantial impact on the budget process" (2003, 31). I have argued here that the fundamental problem of PART, as of all such measures, is that it seeks to analyze programs' overall efficiency so as to compare programs. Such comparisons are not useful to and therefore are not used by Congress—because they inevitably involve value judgments that legislators will not accept from the OMB, and

because the kinds of measures and expertise required are extremely easy to question. In the case of PART, the outcome measures were (and had to be) quite different between programs; other questions were applied even if not obviously appropriate to programs; metrics varied greatly on fundamental factors such as whether they actually measured outcomes; and the standards were chosen by individual examiners in an obscure and discretionary process. This would have reduced credibility even if the examiners had had better claims to expertise. Legislative budgeters are far more interested in information at a more detailed (so less conflictual and value-laden) level that they can use to say, "No, you don't need that much money."

In its first budget proposal, the Barack Obama administration promised to "fundamentally reconfigure" PART while promising to replace "ideological performance goals" with "goals that Americans care about and that are based on congressional intent and feedback from the people served by government programs" (Newell 2009). This responded to Democrats' distrust of the Bush administration without explicitly recognizing the fundamental difficulties with PART. In practice, the process of reconfiguring PART turned into suspending the process. The administration instead chose to emphasize identification and measurement of a small set of "high priority performance goals" that were identified in the *Analytical Perspectives* volume of the president's fiscal year 2011 budget. This approach implicitly recognizes that a central agency must ration its attention to performance issues, while also implicitly abandoning the concept that measures should be used to assess programs overall in a way that might allow comparison.

The retreat from PART's basic logic continued in a memorandum issued on June 25, 2010, in which Office of Personnel Management associate director for performance and personnel management Shelley Metzenbaum (whose research on PART and GPRA I have cited) instructed agencies that they should,

integrate their former Program Assessment Rating Tool (PART) performance measures into their annual performance reports. If an agency wants to eliminate PART or GPRA measures it feels are not useful, it should propose the list of measures for deletion and discuss their deletion with key stakeholders, Congress, and their OMB Resource Management Office before dropping the measures. The list of dropped measures and reasons for their deletion should be included in the annual performance report. (OMB 2010, 5)

Considering the evidence, these are reasonable steps.

It would be helpful, however, if the administration and public administration experts would explain why PART did not work rather than, all too often, raising excessive hopes for what performance measurement could accomplish. While deemphasizing PART is useful in itself, it will not redress the OMB's resource shortages. The OMB needs more staff, and a better ability to retain that staff, to do traditional budgeting better.

Acknowledgments

This article uses unattributed quotations from open-ended interviews with participants in federal budgeting, all of which were conducted on condition that I would neither quote nor attribute views

to any individual without direct written permission. In addition to the interviews conducted in 2008, my understanding of the role of the president's budget within federal budgeting has been influenced by interviews conducted in 1989 and 1990 with 22 OMB staff and 11 departmental or agency budget officers. Since 1983, I have conducted close to 200 other interviews, largely about the broader budget and appropriations processes. About a quarter of these have been with members of Congress; the rest were with a mix of executive branch officials, interest group participants, and, especially, congressional staff. I am exceedingly grateful to all of those who have lent their time and wisdom to my work, and I must emphasize that none of these generous respondents bears any responsibility for errors of fact or judgment in this work.

Notes

1. I use the past tense here because some of the respondents had left OMB at the time of the interviews. Because part of my focus was change over time, I began by contacting 12 individuals whom I had interviewed for previous work; as in a "snowball" sample, I then was directed to others by the original respondents. While I cannot claim that the group was "representative" in any sense, it offered a great deal of expert experience.
2. The OMB examiners also could define a program as "mixed." Here are some examples: the Commodity Supplemental Food Program is a block grant; the National Forest Improvement and Maintenance is capital assets; the Cochran Fellowship Program is a competitive grant; the Rural Water and Wastewater Grants and Loans are credit programs; the Agricultural Commodity Grading and Certification Programs are direct; the Railroad Safety Program is regulatory; the National Institute of Standards and Technology Laboratories are research and development; the Internal Revenue Service's tax collection is mixed.
3. Author's calculations from the "detailed information" on the assessments of each of the programs listed in note, downloaded from <http://www.whitehouse.gov/omb/expectmore> (accessed August 3, 2009).
4. Shea claims that the Bush administration's management efforts improved agency performance but provides no evidence that these improvements were caused by PART. For instance, one might expect that the Social Security Administration would have "increased (its) use of information technology" (Shea 2008, 63) between 2001 and 2008 with or without PART. Gilmour's interview data provide a more reliable test.
5. For powerful critiques of William Niskanen's "budget maximizing bureaucrat" theory, see Blais and Dion (1991) and Wilson (1989).
6. For other examples of appropriations subcommittees expressing distaste for PART, see Frisco and Stalebrink (2009, 13–15).
7. As Stone (1997) emphasizes, "efficiency" in this sense can be neutral only if all parties are concerned with the same output, which may not be the case.

References

- Berman, Larry. 1979. *The Office of Management and Budget and the Presidency, 1921–1979*. Princeton, NJ: Princeton University Press.
- Blais, André, and Stéphane Dion. 1991. *The Budget-Maximizing Bureaucrat: Appraisals and Evidence*. Pittsburgh, PA: University of Pittsburgh Press.
- Blanchard, Lloyd A. 2008. PART and Performance Budgeting Effectiveness. In *Performance Management and Budgeting: How Governments Learn from Experience*, edited by F. Stevens Redburn, Robert J. Shea, and Terry F. Buss, 67–91. Armonk, NY: M. E. Sharpe.
- Blondal, Jon R., Dirk-Jan Kraan, and Michael Ruffner. 2003. Budgeting in the United States. *OECD Journal on Budgeting* 2(2): 7–53.
- Breul, Jonathan D. 2007. Three Bush Administration Management Initiatives: The President's Management Agenda, Freedom to Manage Legislative Proposals, and the Program Assessment Rating Tool. *Public Administration Review* 67(1): 21–26.

Caiden, Naomi. 1987. Paradox, Ambiguity, and Enigma: The Strange Case of the Executive Budget and the United States Constitution. *Public Administration Review* 47(1): 84–92.

Dickinson, Matthew J., and Andrew Rudalevige. 2007. Worked Out in Fractions: Neutral Competence, FDR, and the Bureau of the Budget. *Congress and the Presidency* 34(1): 1–27.

Dull, Matthew. 2006. Why PART? The Institutional Politics of Presidential Budget Reform. *Journal of Public Administration Research and Theory* 16(2): 187–215.

Frederickson, David G., and H. George Frederickson. 2006. *Measuring the Performance of the Hollow State*. Washington, DC: Georgetown University Press.

Frisco, Velda, and Odd J. Stalebrink. 2008. Congressional Use of the Program Assessment Rating Tool. *Public Budgeting and Finance* 28(2): 1–19.

Gilmour, John B. 2007. Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrating Budget and Performance. *OECD Journal on Budgeting* 7(1): 1–39.

———. 2008. Implementing OMB's Program Assessment Rating Tool: Meeting the Challenges of Performance-Based Budgeting. In *Performance Management and Budgeting: How Governments Learn from Experience*, edited by F. Stevens Redburn, Robert J. Shea, and Terry F. Buss, 21–48. Armonk, NY: M. E. Sharpe.

Gilmour, John B., and David E. Lewis. 2006. Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores. *Public Administration Review* 66(5): 742–52.

Greitens, Thomas J., and M. Ernita Joaquin. 2010. Policy Typology and Performance Measurement: Results from the Program Assessment Rating Tool (PART). *Public Performance and Management Review* 33(4): 555–70.

Gueorguieva, Vassia, Jean Accius, Carmen Apaza, Lamar Bennett, Clinton Brownley, Shea Cronin, and Panote Preechyanud. 2009. The Program Assessment Rating Tool and the Government Performance and Results Act: Evaluating Conflicts and Disconnections. *American Review of Public Administration* 39(3): 225–45.

Hecl, Hugh. 1984. Executive Budget Making. In *Federal Budget Policy in the 1980s*, edited by Gregory B. Mills and John L. Palmer, 225–91. Washington, DC: Urban Institute.

Johnson, Bruce E. 1984. From Analyst to Negotiator: The OMB's New Role. *Journal of Policy Analysis and Management* 3(4): 501–15.

Kamensky, John. 2007. Federal Agencies Ratchet Up Focus on Performance. *PA Times*, April, 7.

Kasdin, Stuart. 2010. Reinventing Reforms: How to Improve Program Management Using Performance Measures. Really. *Public Budgeting and Finance* 30(3): 51–78.

Matthiasen, David G. 1988. The Evolution of the Office of Management and Budget under President Reagan. *Public Budgeting and Finance* 8(3): 3–14.

McMurtry, Virginia A. 2005. *Performance Management and Budgeting in the Federal Government: Brief History and Recent Developments*. Report no. RL32164. Washington, DC: Congressional Research Service.

Metzenbaum, Shelley H. 2009. *Performance Management Recommendations for New Administration*. Washington, DC: IBM Center for the Business of Government.

Mosher, Frederick C. 1984. *A Tale of Two Agencies: A Comparative Analysis of the General Accounting Office and the Office of Management and Budget*. Baton Rouge: Louisiana State University Press.

Nathan, Richard P. 2005. Presidential Address: "Complexifying" Performance Oversight in America's Governments. *Journal of Policy Analysis and Management* 24(3): 207–15.

———. 2008. The Role of Performance Management in American Federalism. In *Performance Management and Budgeting: How Governments Learn from Experience*, edited by F. Stevens Redburn, Robert J. Shea, and Terry F. Buss, 170–77. Armonk, NY: M. E. Sharpe.

Newell, Elizabeth. 2009. Budget Summary Is Littered with Management Proposals. *Government Executive*, February 26.

Posner, Paul L. 2007. The Continuity of Change: Public Budgeting and Finance Reforms over 70 Years. *Public Administration Review* 67(6): 1020–29.

Posner, Paul L., and Denise M. Fantone. 2008. Performance Budgeting: Prospects for Sustainability. In *Performance Management and Budgeting: How Governments Learn from Experience*, edited by F. Stevens Redburn, Robert J. Shea, and Terry F. Buss, 92–113. Armonk, NY: M. E. Sharpe.

Radin, Beryl A. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.

———. 2008. The Legacy of Federal Management Change: PART Repeats Familiar Problems. In *Performance Management and Budgeting: How Governments Learn from Experience*, edited by F. Stevens Redburn, Robert J. Shea, and Terry F. Buss, 114–32. Armonk, NY: M. E. Sharpe.

Redburn, F. Stevens, Robert J. Shea, Terry F. Buss, and Ednilson Quintanilla. 2008. Performance-Based Management: How Governments Can Learn from Experience. In *Performance Management and Budgeting: How Governments Learn from Experience*, edited by F. Stevens Redburn, Robert J. Shea, and Terry F. Buss, 3–18. Armonk, NY: M. E. Sharpe.

Rubin, Irene. 1994. Early Budget Reformers: Democracy, Efficiency, and Budget Reforms. *American Review of Public Administration* 24(3): 229–52.

Schick, Allen. 1966. The Road to PPB. *Public Administration Review* 26(4): 243–58.

———. 1973. A Death in the Bureaucracy: The Demise of Federal PPB. *Public Administration Review* 33(2): 146–56.

———. 1980. *Congress and Money: Budgeting, Spending, and Taxing*. Washington, DC: Urban Institute.

———. 2004. Twenty-Five Years of Budgeting Reform. *OECD Journal on Budgeting* 4(1): 81–102.

Shea, Robert J. 2008. Performance Budgeting in the United States. *OECD Journal on Budgeting* 8(1): 61–73.

Simon, Herbert A. 1997. *Administrative Behavior*. 4th ed. New York: Free Press.

Skelly, Tom. 2006. Improving Integration of Performance Information in Congressional Budget Justifications. Presentation to the Meeting of the Budget and Performance Integration Leads, November 8. <http://georgewebush-whitehouse.archives.gov/results/agenda/fy06q4-bpiupdate.html> [accessed September 15, 2011].

Stalebrink, Odd J. 2009. National Performance Mandates and Intergovernmental Collaboration: An Examination of the Program Assessment Rating Tool (PART). *American Review of Public Administration* 39(6): 619–39.

Stone, Deborah. 1997. *Policy Paradox: The Art of Political Decision Making*. 2nd ed. New York: W. W. Norton.

Stone, Donald. 1990. Administrative Management: Reflections on Origins and Accomplishments. *Public Administration Review* 50(1): 3–20.

Tanaka, Susan, June O' Neill, and Arlene Holen. 2003. "Above the Fray": The Role of the U.S. Office of Management and Budget. In *Controlling Public Expenditure: The Changing Roles of Central Budget Agencies—Better Guardians?* edited by John Wanna, Lotte Jensen, and Jouke de Vries, 57–84. Northampton, MA: Edward Elgar.

Thompson, Frank. 2002. Reinvention in the States: Ripple or Tide? *Public Administration Review* 62(3): 362–67.

Tomkin, Shelley Lynne. 1998. *Inside OMB: Politics and Process in the President's Budget Office*. Armonk, NY: M. E. Sharpe.

U.S. General Accounting Office (GAO). 1997. *The Results Act: Observations on the Office of Management and Budget's July 1997 Draft Strategic Plan*. Washington, DC: U.S. Government Printing Office. GAO/AIMD/GGD/97-169R.

U.S. Government Accountability Office (GAO). 2008. *Government Performance: Lessons Learned for the Next Administration on Using Performance Information to Improve Results*. Washington, DC: U.S. Government Printing Office. GAO-08-1026T.

U.S. House of Representatives. 2008. Committee on Appropriations. House Report 110-207: Financial Services and General Government Appropriations Bill, 2008. <http://www.gpo.gov/fdsys/plg/CRPT-110hrpt207/html/CRPT-110hrpt207.htm> [accessed September 15, 2011].

U.S. Office of Management and Budget (OMB). 2003. The Budget of the United States Government, Fiscal Year 2004, Rating the Performance of Federal Programs. <http://www.gpoaccess.gov/usbudget/fy04/pdf/budget/performance.pdf> [accessed September 15, 2011].

———. 2010. Memorandum Re: Performance Improvement Guidance: Management Responsibilities and Government Performance and Results Act Documents. June 25. http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-24.pdf [accessed September 15, 2011].

U.S. Senate. 2008. *Committee on Appropriations. Committee on Appropriations U.S. Senate: 1867–2008*. 110th Cong., 2nd sess., Doc. no. 14. Washington, DC: U.S. Government Printing Office.

White, Joseph. 1991. Presidential Power and the Budget. In *Federal Budget and Financial Management Reform*, edited by Thomas D. Lynch, 1–29. New York: Quorum.

———. 2009. The President's Budget vs. Congressional Budgeting: Institutionalizing the Adversarial Presidency? In *Rivals for Power: Presidential–Congressional Relations*, 4th ed., edited by James A. Thurber, 229–49. Lanham, MD: Rowman & Littlefield.

White, Joseph, and Aaron Wildavsky. 1991. *The Deficit and the Public Interest: The Search for Responsible Budgeting in the 1980s*. Berkeley: University of California Press.

Wichowsky, Amber, and Donald P. Moynihan. 2008. Measuring How Administration Shapes Citizenship: A Policy Feedback Perspective on Performance Management. *Public Administration Review* 68(5): 908–20.

Wildavsky, Aaron. 1964. *The Politics of the Budgetary Process*. Boston: Little, Brown.

———. 1966. The Political Economy of Efficiency: Cost-Benefit Analysis, Systems Analysis, and Program Budgeting. *Public Administration Review* 26(4): 292–310.

———. 1969. Rescuing Policy Analysis from PPBS. *Public Administration Review* 29(2): 189–202.

———. 1981. Budgetary Reform in an Age of Big Government. In *Contemporary Public Administration*, edited by Thomas Vocino and Jack Rabin, 261–91. New York: Harcourt, Brace, Jovanovich.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Barry Anderson
National Governors Association

Commentary: "Playing the Wrong PART: The Program Assessment Rating Tool and the Functions of the President's Budget"

The extensive research that I and others did at the Organisation for Economic Co-operation and Development and at the International Monetary Fund indicates that countries—and the analysts who review them—often confuse performance budgeting with performance-informed budgeting. The former is much more involved with the allocation of resources based on achieved results. In the latter, performance information is important in the budget decision-making process, but it does not determine the amount of resources allocated, nor does it have a predefined weight in decisions. This confusion is evident in the application of Program Assessment Rating Tool (PART) information in the United States, as described in Professor Joseph White's article.

PART is the latest in a series of efforts that began more than 60 years ago to develop and use

performance information in the U.S. budget process. To the extent that PART information was intended to be used directly in the allocation of resources, then, for the reasons pointed out in White's article, I concur with the conclusion that "PART did not help." However, if the information produced by PART was to be considered in the decision-making process, but certainly not as the only or necessarily the determining factor in resource allocation, then PART information could contribute to the process. In this context, it is important not to underestimate the value of efforts—even generally unsuccessful ones—to reorient the budget decision process away from inputs, such as appropriation amounts or staffing levels, and toward outputs and eventually, in some cases, outcomes. PART's proponents may well have promised much more than it could deliver and, in the process, consumed staff resources that could have been better

Joseph White is the Luxenberg Family Professor of Public Policy in the Department of Political Science at Case Western Reserve University. His research focuses on federal budget policy and politics; health care policy and politics in the United States and other advanced industrial democracies; policy consequences of an aging society; and policies to control health care costs. He is the coauthor with Aaron Wildavsky of *The Deficit and the Public Interest: The Search for Responsible Budgeting in the 1980s*.
E-mail: joseph.white@case.edu

Public Administration Review,
Vol. 72, Iss. 1, pp. 121–122. © 2011 by
The American Society for Public Administration.
DOI: 10.1111/j.1540-6210.2011.02495.x

utilized on more direct output measures. But PART's goals are hard to fault even if its execution left much to be desired.

The international budget community has closely followed, and learned from, the U.S. experiences with PART. In particular, PART's management scorecard—red, yellow, and green scores or, using its politically incorrect label, “name and shame”—has been copied. Perhaps many experienced and oh so savvy U.S. officials did indeed learn “how to take the [ratings] test” without creating “a culture of managing for performance.” But no other country is committed to a separation of powers mentality like the United States, and thus the power of name and shame and the information produced to derive a scorecard has had much

greater resonance overseas. Even in the United States, the cynical assessment of many of PART's value may be as much a comment on the widely discredited fiscal policies of President George W. Bush as on the value and use of PART itself. Nevertheless, there is no doubt that PART could have contributed much more had it not raised “excessive hopes for what performance measurement could accomplish.”

The United States produces more and better-quality information on performance and results than any other country. However, it will not take advantage of this information unless it is integrated into a performance-informed budgeting system, not a performance budgeting one. It is hoped the PART experience will help the United States reach this conclusion.

How Credible Is the Evidence, and Does It Matter? An Analysis of the Program Assessment Rating Tool

This research empirically assesses the quality of evidence that agencies provided to the Office of Management and Budget in the application of the Program Assessment Rating Tool (PART), introduced in 2002 to more rigorously, systematically, and transparently assess public program effectiveness and hold agencies accountable for results by tying them to the executive budget formulation process and program funding. Evidence submitted by 95 programs administered by the U.S. Department of Health and Human Services for the PART assessment is analyzed using measures that capture the quality of evidence and methods used by programs and information on characteristics of agencies that might relate to program results and government funding decisions. The study finds that of those programs offering some evidence, most was internal and qualitative, and about half did not assess how their performance compared to other government or private programs with similar objectives. Programs were least likely to provide externally generated evidence of their performance relative to long-term and annual performance goals. Importantly, overall PART and results scores were (statistically) significantly lower for programs that failed to provide quantitative evidence and did not use long-term measures, baseline measures or targets, or independent evaluations. Although the PART program results ratings and overall PART scores had no discernible consequences for program funding over time, the PART assessments appeared to take seriously the evaluation of evidence quality, a positive step forward in recent efforts to base policy decisions on more rigorous evidence.

In the 1990s, “results-oriented government” took off as a new way of holding government accountable for how it spends public money and, in particular, for the outcomes or results it produces. The new tools and public management reforms advanced reflected an intentional shift from an emphasis on rules- or compliance-oriented accountability toward a focus on performance, or how well an organization does what it does in relation to its organizational goals (Heinrich 2003; Radin 2000). Although prior administrations initiated reforms promoting accountability for results, pay for performance,

and performance-based contracting, the National Partnership for Reinventing Government, spearheaded by Vice President Al Gore (and drawing on the influential work of Osborne and Gaebler 1992), transformed these themes and principles into a movement to improve government performance, complete with reinvention teams (internal and government-wide), reinvention laboratories within agencies, town hall meetings and reinvention summits, and new legislation to mandate performance management at the federal level (Kamensky 1999).¹

The first major fruit of these efforts was the Government Performance and Results Act (GPRA), enacted in 1993 to generate more objective information on government performance and efficiency by measuring progress toward performance goals, providing evidence of performance relative to targets, and holding federal agencies accountable for results in annual reports to the public. In the decade and a half since passage of the GPRA, few dispute that there has been a definitive transformation in federal government capacity and infrastructure for managing for results, that is, in its use of outcome-oriented strategic plans, performance measures, and reporting of results (GAO 2008). Yet some in-depth assessments of the implementation of the GPRA also have been highly critical. Several researchers have suggested that overlaying a results-oriented managerial logic on top of an inherently political process in which agency goals may be ambiguous or contradictory sets the stage for inevitable problems in implementation, above and beyond the challenges of identifying adequate measures of performance (Frederickson and Frederickson 2006; Radin 2000, 2006). Radin (2000) argued that rather than freeing public managers to focus on results, the GPRA's performance requirements exacerbated administrative constraints and conflict among program managers and heightened distrust between agencies and legislators.

One of the primary goals of the George W. Bush administration in introducing the Program

Carolyn J. Heinrich
University of Texas at Austin

Carolyn J. Heinrich is the Sid Richardson Professor of Public Affairs and director of the Center for Health and Social Policy in the Lyndon B. Johnson School of Public Affairs at the University of Texas at Austin. She previously served as director and professor in the La Follette School of Public Affairs at the University of Wisconsin–Madison and associate director of the Institute for Research on Poverty and assistant professor at the University of North Carolina at Chapel Hill. She holds a doctorate in public policy studies from the University of Chicago. E-mail: cheinrich@austin.utexas.edu