

CRS Report for Congress

Received through the CRS Web

The Bush Administration's Program Assessment Rating Tool (PART)

November 5, 2004

Clinton T. Brass
Analyst in American National Government
Government and Finance Division

The Bush Administration's Program Assessment Rating Tool (PART)

Summary

Federal government agencies and programs work to accomplish widely varying missions. These agencies and programs employ a number of public policy approaches, including federal spending, tax laws, tax expenditures, and regulation. Given the scope and complexity of these efforts, it is understandable that citizens, their elected representatives, civil servants, and the public at large would have an interest in the performance and results of government agencies and programs.

Evaluating the performance of government agencies and programs has proven difficult and often controversial. In spite of these challenges, in the last 50 years both Congress and the President have undertaken numerous efforts — sometimes referred to as performance management, performance budgeting, strategic planning, or program evaluation — to analyze and manage the federal government's performance. Many of those initiatives attempted in varying ways to use performance information to influence budget and management decisions for agencies and programs. The George W. Bush Administration's release of the Program Assessment Rating Tool (PART) is the latest of these efforts.

The PART is a set of questionnaires that the Bush Administration developed to assess the effectiveness of different types of federal executive branch programs, in order to influence funding and management decisions. A component of the President's Management Agenda (PMA), the PART focuses on four aspects of a program: purpose and design; strategic planning; program management; and program results/accountability. The Administration submitted PART ratings for programs along with the President's FY2004 and FY2005 budget proposals, and plans to continue doing so for FY2006 and subsequent years.

This report discusses how the PART is structured, how it has been used, and how various commentators have assessed its design and implementation. The report concludes with a discussion of potential criteria for assessing the PART or other program evaluations, which Congress might consider during the budget process, in oversight of federal agencies and programs, and in consideration of legislation that relates to the PART or program evaluation generally.

Proponents have seen the PART as a necessary enhancement to the Government Performance and Results Act (GPRA), a law that the Administration views as not having met its objectives, in order to hold agencies accountable for performance and to integrate budgeting with performance. However, critics have seen the PART as overly political and a tool to shift power from Congress to the President, as well as failing to provide for adequate stakeholder consultation and public participation. Some observers have commented that the PART has provided a needed stimulus to agency program evaluation efforts, but they do not agree on whether the PART validly assesses program effectiveness.

This report will be updated as events warrant.

Contents

Introduction	1
Development and Use of the PART	3
OMB on the PART's Purpose	3
PART Components and Process	5
Publication and Presentation	7
OMB Statements About the PART	8
Transparency and Objectivity	8
OMB Use of PART Ratings	10
Third-Party Assessments of the PART	11
Performance Institute	11
Scholarly Assessments	12
The PART	12
PART and Performance Budgeting	14
Government Accountability Office	15
Potential Criteria for Evaluating the PART or Other Program Evaluations	18
Concepts for Evaluating a Program Evaluation	18
Evaluating the PART	20

List of Tables

Table 1. OMB Rating Categories for the PART	8
---	---

The Bush Administration's Program Assessment Rating Tool (PART)

The Program Assessment Rating Tool (PART) is a set of questionnaires that the George W. Bush Administration developed to assess the effectiveness of different types of federal executive branch programs, in order to influence funding and management decisions. A component of the President's Management Agenda, the PART focuses on four aspects of a program: purpose and design; strategic planning; program management; and program results/accountability. In July 2002, the Office of Management and Budget (OMB) issued the PART to executive branch agencies for use in the upcoming FY2004 budget cycle to assess programs representing approximately 20% of the federal budget. For succeeding budget cycles, OMB said that additional 20% increments of federal programs would be assessed with the PART. The Administration subsequently released PART ratings for selected programs along with the President's FY2004 and FY2005 budget proposals, which were transmitted to Congress on February 3, 2003, and February 2, 2004, respectively. Another round of ratings is planned to be released with the President's FY2006 budget.

This report discusses how the PART is structured, how it has been used, and how various commentators have assessed its design and implementation. Finally, the report discusses potential criteria for assessing the PART, or other program evaluations, which Congress might consider during the budget process, in oversight of federal agencies and programs, and in consideration of legislation that relates to the PART or program evaluation generally.

Introduction

Federal government agencies and programs work to accomplish widely varying missions. These agencies and programs use a number of public policy approaches, including federal spending, tax laws, tax expenditures,¹ and regulation.² In FY2004,

¹ *Tax expenditures* are generally defined as revenue losses (reductions in tax liabilities) from preferential provisions in tax laws. For more information, see CRS Electronic Briefing Book, *Taxation*, page on "Tax Expenditures," by Jane G. Gravelle, at [<http://www.congress.gov/brbk/html/ebtxr9.html>].

² Data on federal outlays are available in U.S. Congressional Budget Office, *The Budget and Economic Outlook: An Update* (Washington: CBO, Sept. 2004), p. x. Tax expenditure data are available in U.S. Office of Management and Budget, *Budget of the U.S. Government, Fiscal Year 2005, Analytical Perspectives* (Washington: GPO, 2004), pp. 296-299. For a comparison of the relative size of tax expenditures with federal outlays, see CRS Report (continued...)

estimated federal spending was \$2.3 trillion, and tax expenditures totaled approximately \$1 trillion. Estimates of the off-budget costs of federal regulations have ranged in the hundreds of billions of dollars, and corresponding estimates of benefits of federal regulations have ranged from the hundreds of billions to trillions of dollars.³

Given the scope and complexity of these various efforts, it is understandable that citizens, their elected representatives, civil servants, and the public at large would have an interest in the performance and results of government activities. Evaluating the performance of government agencies and programs, however, has proven difficult and often controversial:

- Actors in the U.S. political system (e.g., Members of Congress, the President, citizens, interest groups) often disagree about the appropriate uses of public funds; missions, goals, and objectives for public programs; and criteria for evaluating success. One person's key program may be another person's key example of waste and abuse, and different people have different conceptions of what "good performance" means.
- Even when consensus is reached on a program's appropriate goals and evaluation criteria, it is often difficult and sometimes almost impossible to separate the discrete influence that a federal program had on key outcomes from the influence of other actors (e.g., state and local governments), trends (e.g., globalization, demographic changes), and events (e.g., natural disasters).
- Federal agencies and programs often have multiple purposes, and sometimes these purposes may conflict or be in tension with one another. Finding and assessing a balance among priorities can be controversial and difficult.
- The outcomes of some agencies and programs are viewed by many observers as inherently difficult to measure. Foreign policy and research and development programs have been cited as examples.
- There is frequently a time lag between an agency's or program's actions and eventual results (or lack thereof). In the absence of this eventual outcome data, it is often difficult to know how to assess if a program is succeeding.

² (...continued)

RS21710, *Tax Expenditures Compared with Outlays by Budget Function: Fact Sheet*, by Nonna A. Noto.

³ For analysis of benefit and cost estimates of regulations, see CRS Report RL32339, *Federal Regulations: Efforts to Estimate Total Costs and Benefits of Rules*, by Curtis W. Copeland.

- Many observers have asserted that agencies do not adequately evaluate the performance or results of their programs — or integrate evaluation efforts across agency boundaries — possibly due to lack of capacity, management attention and commitment, or resources.⁴

In spite of these and other challenges,⁵ in the last 50 years both Congress and the President have undertaken numerous efforts — sometimes referred to as performance management, performance budgeting, strategic planning, or program evaluation — to analyze and manage the federal government’s performance. Many of those initiatives attempted in varying ways to use performance information to influence budget and management decisions for agencies and programs.⁶ The Bush Administration’s release of PART ratings along with the President’s FY2004 and FY2005 budget proposals, and its plans to continue doing so for FY2006 and subsequent years, represent the latest of these efforts.

Development and Use of the PART

OMB on the PART’s Purpose

The PART was created by OMB within the context of the Bush Administration’s broader Budget and Performance Integration (BPI) initiative, one of five government-wide initiatives under the President’s Management Agenda (PMA).⁷ According to the President’s proposed FY2005 budget, the goal of the BPI

⁴ For example, see the General Accounting Office testimony in U.S. Congress, House Committee on Government Reform, Subcommittee on Government Efficiency and Financial Management, *Performance, Results, and Budget Decisions*, hearing, 108th Cong., 1st sess., Apr. 1, 2003, H.Hrg. 108-32 (Washington: GPO, 2003), pp. 30-31. For historical context, see U.S. General Accounting Office, *Transition Series: Program Evaluation Issues*, GAO/OCG-93-6TR, Dec. 1992; and Walter Williams, *Mismanaging America: The Rise of the Anti-Analytic Presidency* (Lawrence, KS: University Press of Kansas, 1990). The General Accounting Office is now called the Government Accountability Office, but this report will use the previous name when citing sources published under that name.

⁵ For a similar listing of challenges, see U.S. Office of Management and Budget, “Performance Measurement Challenges and Strategies,” June 18, 2003, available at [http://www.whitehouse.gov/omb/part/challenges_strategies.pdf]. See also Graham T. Allison Jr., “Public and Private Management: Are They Fundamentally Alike in All Unimportant Respects?,” in Frederick S. Lane, *Current Issues in Public Administration*, 2nd ed. (New York: St. Martin’s Press, 1982), p. 18; Peter F. Drucker, *Management* (New York: Harper & Row, 1974), pp. 130-166; and Henry Mintzberg, “Managing Government, Governing Management,” *Harvard Business Review*, May-June 1996, pp. 79-80.

⁶ For a brief history and summary of developments in the areas of performance management and budgeting, including the PART, see CRS Report RL32164, *Performance Management and Budgeting in the Federal Government: Brief History and Recent Developments*, by Virginia A. McMurtry.

⁷ For an overview of OMB’s organization and functions, see CRS Report RS21665, *Office of Management and Budget: A Brief Overview*, by Clinton T. Brass. For an overview of the (continued...)

initiative is to “have the Congress and the Executive Branch routinely consider performance information, among other factors, when making management and funding decisions.”⁸ In turn,

[the PART] is designed to help assess the management and performance of individual programs. The PART helps evaluate a program’s purpose, design, planning, management, results, and accountability to determine its ultimate effectiveness.⁹

The PART evaluates executive branch programs that have funding associated with them.¹⁰ The Bush Administration submitted approximately 400 PART scores and analyses along with the President’s FY2004 and FY2005 budget proposals,¹¹ with the intent to assess programs amounting to approximately 20% of the federal budget each fiscal year for five years, from FY2004 to FY2008. For FY2004, OMB assessed 234 programs. For FY2005, a further 173 programs were assessed.¹² For these two years combined, OMB said that about 40% of the federal budget, or nearly \$1.1 trillion, had been “PARTed.”

In releasing the PART, the Bush Administration asserted that Congress’s current statutory framework for executive branch strategic planning and performance reporting, the Government Performance and Results Act of 1993 (GPRA),

[w]hile well-intentioned ... did not meet its objectives. Through the President’s Budget and Performance Integration initiative, augmented by the PART, the Administration will strive to implement the objectives of GPRA.¹³

⁷ (...continued)

PMA, see CRS Report RS21416, *The President’s Management Agenda: A Brief Introduction*, by Virginia A. McMurtry.

⁸ U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2005, Analytical Perspectives* (Washington: GPO, 2004), p. 9.

⁹ *Ibid.*

¹⁰ The PART does not formally assess policy tools that are not the subject of appropriations, such as stand-alone tax expenditures or tax laws.

¹¹ For FY2004 PART-related information, see U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2004, Performance and Management Assessments* (Washington: GPO, 2003), and OMB’s website at [<http://www.whitehouse.gov/omb/budget/fy2004/pma.html>], which includes each program’s PART “Summary” (in PDF format) and detailed “Worksheet” (in Microsoft Excel format). For FY2005 PART-related information, see U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2005, Analytical Perspectives*, pp. 7-22, along with accompanying CD-ROM for PART “Program Summary” and data files (electronic PDF files and single Microsoft Excel file); and OMB’s website at [<http://www.whitehouse.gov/omb/budget/fy2005/part.html>], which includes each program’s detailed PART worksheet (in PDF).

¹² Some programs that were assessed for FY2004 were reassessed for FY2005. In addition, some programs that were assessed individually for FY2004 were combined into single programs for presentation in the President’s FY2005 budget.

¹³ U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal* (continued...)

As discussed later in this report, this move and the PART's perceived lack of integration with GPRA was controversial among some observers, in part because OMB, and by extension the Bush Administration, were seen as "substituting [their] judgment" about agency strategic planning and program evaluations "for a wide range of stakeholder interests" under the framework established by Congress under GPRA.¹⁴ Under GPRA, 5 U.S.C. § 306 requires an agency when developing its strategic plan¹⁵ to consult with Congress and "solicit and consider the views and suggestions of those entities potentially affected by or interested in such a plan." Some observers have recommended a stronger integration between PART and GPRA, thereby more strongly integrating executive and congressional management reform efforts.¹⁶

PART Components and Process

OMB developed seven versions of the PART questionnaire for different types of programs.¹⁷ Structurally, each version of the PART has approximately 30 questions that are divided into four sections. Depending on how the questionnaire is filled in and evaluated, each section provides a percentage "effectiveness" rating (e.g., 85%). The four sections are then averaged to create a single PART score according to the following weights: (1) program purpose and design, 20%; (2) strategic planning, 10%; (3) program management, 20%; and (4) program results/accountability, 50%.

¹³ (...continued)

Year 2004, Performance and Management Assessments, p. 9.

¹⁴ Quotes from U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, GAO-04-174, Jan. 2004, *Highlights*, inside front cover. For an overview of GPRA, see CRS Report RL30795, *General Management Laws: A Compendium*, entry for "Government Performance and Results Act of 1993" in section II.B. of the report, by Genevieve J. Knezo.

¹⁵ Under GPRA, an agency strategic plan specifies goals and objectives, the relationship between those goals and objectives and the agency's annual performance plan, and program evaluations, among other things.

¹⁶ For related comments from OMB's Performance Measurement Advisory Council (PMAC), which provided early input to OMB on the draft PART, see [http://www.whitehouse.gov/omb/budintegration/pmac_index.html], and specifically [http://www.whitehouse.gov/omb/budintegration/pmac_030303comments.pdf]. See also U.S. General Accounting Office, *Performance Budgeting: OMB's Program Assessment Rating Tool Presents Opportunities and Challenges for Budget and Performance Integration*, GAO-04-439T, Feb. 4, 2004.

¹⁷ These "program types" include direct federal programs; competitive grant programs; block/formula grant programs; regulatory based programs; capital assets and service acquisition programs; credit programs; and research and development programs. For OMB's definitions of these types, see U.S. Office of Management and Budget, *Instructions for the Program Assessment Rating Tool* (undated), p. 12, available at OMB's website ([<http://www.whitehouse.gov/omb/part/index.html>]) under the link "PART Guidance for FY2006 Budget," at [http://www.whitehouse.gov/omb/part/2006_part_guidance.pdf].

Under the overall supervision of OMB and agency political appointees, OMB's program examiners and agency staff negotiate and complete the questionnaire for each "program" — thereby determining a program's section and overall PART scores. In the event of disagreements between OMB and agencies regarding PART assessments, OMB's PART instructions for FY2005 stated that "[a]greements on PART scoring should be reached in a manner consistent with settling appeals on budget matters."¹⁸ Under that process, scores are ultimately decided or approved by OMB political appointees and the White House. When the PART questionnaire responses are completed, agency and OMB staff prepare materials for inclusion in the President's annual budget proposal to Congress.

According to OMB's most recent guidance to agencies for the PART, the definition of *program* will most often be determined by a budgetary perspective. That is, the "program" that OMB assesses with the PART will most often be what OMB calls a *program activity*, or aggregation of *program activities*, as listed in the President's budget proposal:

One feature of the PART process is flexibility for OMB and agencies to determine the unit of analysis — "program" — for PART review. The structure that is readily available for this purpose is the formal budget structure of accounts and activities supporting budget development for the Executive Branch and the Congress and, in particular, Congressional appropriations committees.... Although the budget structure is not perfect for program review in every instance — for example, "program activities" in the budget are not always the activities that are managed as a program in practice — the budget structure is the most readily available and comprehensive system for conveying PART results transparently to interested parties throughout the Executive and Legislative Branches, as well as to the public at large.¹⁹

The term *program activity* is essentially defined by OMB's *Circular No. A-11* as the activities and projects financed by a budget account (or a distinct subset of the activities and projects financed by a budget account), as those activities are outlined in the President's annual budget proposal.²⁰ As noted later, this budget-centered

¹⁸ U.S. Office of Management and Budget, "Completing the Program Assessment Rating Tool (PART) for the FY2005 Review Process," Budget Procedures Memorandum No. 861, p. 4. According to an OMB official who was quoted in a magazine article, about 40% of scores were appealed. See Matthew Weinstock, "Under the Microscope," *Government Executive*, vol. 35, no. 1 (Jan. 2003), p. 39.

¹⁹ Excerpted from U.S. Office of Management and Budget, *Instructions for the Program Assessment Rating Tool*, March 22, 2004, p. 3, available at OMB's website ([<http://www.whitehouse.gov/omb/part/index.html>]) under the link "PART Guidance for FY2006 Budget," at [http://www.whitehouse.gov/omb/part/2006_part_guidance.pdf]. Agencies and OMB decide each year what "programs" will be assessed by the PART.

²⁰ *Circular No. A-11* is OMB's instruction to agencies on how to prepare budget submissions for review within the executive branch and, separately, for eventual presentation to Congress. For each budget account, *Circular No. A-11* instructs agencies to identify one or more program activities according to several general criteria (e.g., keeping the number of program activities to a reasonable minimum without sacrificing clarity, financing no more

(continued...)

approach has been criticized by some observers, because this budget perspective did not necessarily match an agency's organization or strategic planning.

Publication and Presentation

For each program that has been assessed, OMB develops a one-page "Program Summary" that is publicly available in electronic PDF format.²¹ Each summary displays four separate scores, as determined by OMB, for the PART's four sections. OMB also made available for each program a detailed PART "worksheet" to briefly show how each question and section of the questionnaire was filled in, evaluated, and scored.²²

OMB states that the numeric scores for each section are used to generate an overall effectiveness rating for each "program":

[The section scores] are then combined to achieve an overall qualitative rating of either Effective, Moderately Effective, Adequate, or Ineffective. Programs that do not have acceptable performance measures or have not yet collected performance data generally receive a rating of Results Not Demonstrated.²³

The PART's overall "qualitative" rating is ultimately driven by a single numerical score. However, none of OMB's FY2005 budget materials, one-page program summaries, or detailed worksheets displays a program's overall numeric score according to OMB's PART assessment. OMB stated that it does not publish these single numerical scores, because "numerical scores are not so precise as to be able to reliably compare differences of a few points among different programs.... [Overall scores] are rather used as a guide to determine qualitative ratings that are more generally comparable across programs."²⁴ However, these composite weighted scores can be computed manually using OMB's weighting formula.²⁵

²⁰ (...continued)

than one strategic goal or objective, and others). For more information, see OMB *Circular No. A-11*, "Preparation, Submission, and Execution of the Budget," July 2004, Section 82.2, available at [<http://www.whitehouse.gov/omb/circulars/index.html>].

²¹ For FY2005, see [<http://www.whitehouse.gov/omb/budget/fy2005/part.html>], under the "Program summaries" link, or the CD-ROM from OMB's FY2005 *Analytical Perspectives* volume.

²² For FY2005 worksheets, see [<http://www.whitehouse.gov/omb/budget/fy2005/part.html>], under the "PART assessment details" heading, for links to agency-specific PDF files that contain all of each agency's PART worksheets.

²³ U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2005, Analytical Perspectives*, p. 10.

²⁴ See "PART Frequently Asked Question" #28, available at [http://www.whitehouse.gov/omb/part/2004_faq.html].

²⁵ To obtain the spreadsheet file, see OMB's website at [<http://www.whitehouse.gov/omb/budget/fy2005/sheets/part.xls>], or see the CD-ROM from OMB's FY2005 *Analytical Perspectives* volume. When assigning final ratings, OMB used the formula weights to
(continued...)

The only PART effectiveness rating that OMB defines explicitly is “Results Not Demonstrated,” as shown by the excerpt above.²⁶ The Government Accountability Office (GAO, formerly the General Accounting Office) has stated that “[i]t is important for users of the PART information to interpret the ‘results not demonstrated’ designation as ‘unknown effectiveness’ rather than as meaning the program is ‘ineffective.’”²⁷ The other four ratings, which are graduated from best to worst, are driven directly by each program’s overall quantitative score, as outlined in the following table.

Table 1. OMB Rating Categories for the PART

PART “Effectiveness Rating”	Overall Weighted Score
Effective	85% - 100%
Moderately Effective	70% - 84%
Adequate	50% - 69%
Ineffective	0% - 49%
Results Not Demonstrated	n/a (OMB determination that, regardless of score, program measures are unacceptable or that performance data have not yet been collected. This rating does not equate with “ineffective.”)

Source: OMB’s website, at [http://www.whitehouse.gov/omb/part/2004_faq.html], “PART Frequently Asked Question” #29. This website appears to be the only publicly available location where OMB indicates how OMB translated numerical scores into overall “qualitative” ratings.

OMB Statements About the PART

Transparency and Objectivity

OMB has stated that it wants to make the PART process and scores transparent, consistent, systematic, and objective. To that end, OMB solicited and received feedback and informal comments from agencies, congressional staff, GAO, and “outside experts” on ways to change the instrument before it was published with the

²⁵ (...continued)

determine overall scores and appears to have then rounded those figures to the nearest percentage point.

²⁶ OMB further explains on its website that the rating “Results Not Demonstrated” means “that a program does not have sufficient performance measurement or performance information to show results, and therefore it is not possible to assess whether it has achieved its goals.” See “PART Frequently Asked Question” #33, available at [http://www.whitehouse.gov/omb/part/2004_faq.html].

²⁷ See U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB’s Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, p. 25.

President's FY2004 budget proposal in February 2003.²⁸ In an effort to increase transparency, for example, OMB made the detailed PART worksheets available for each program. To make PART assessments more consistent, OMB subjected its assessments to a consistency check.²⁹ That review was "examined," in turn, by the National Academy of Public Administration (NAPA).³⁰ To make the PART more systematic, OMB established formal criteria for assessing programs and created an instrument that differentiated among the seven types of programs (e.g., credit programs, research and development programs).

With regard to the goal of achieving objectivity, OMB made changes to the draft PART before its release in February 2003 with the President's budget. For example, OMB eliminated a draft PART question on whether a program was appropriate at the federal level, because OMB found that question "was too subjective and [assessments] could vary depending on philosophical or political viewpoints."³¹ However, OMB went further to state:

While subjectivity can be minimized, it can never be completely eliminated regardless of the method or tool. In providing advice to OMB Directors, OMB staff have always exercised professional judgment with some degree of subjectivity. That will not change.... [T]he PART makes public and transparent

²⁸ For documents relating to OMB's Performance Measurement Advisory Council (PMAC), which provided early input to OMB on the draft PART, see [http://www.whitehouse.gov/omb/budintegration/pmac_index.html]. For media coverage of the PMAC's input on the draft PART, see Diane Frank, "Council Asks for Tweak of OMB Tool," *FCW.com*, July 8, 2004, available at [<http://www.fcw.com/fcw/articles/2002/0708/pol-omb-07-08-02.asp>]. See also U.S. Office of Management and Budget, "Program Performance Assessments for the FY 2004 Budget," Memorandum for Heads of Executive Departments and Agencies from Mitchell E. Daniels Jr., M-02-10, July 16, 2002, p. 1. This memorandum outlines selected changes to the draft PART instrument before it was used for the President's FY2004 budget proposal; available at [<http://www.whitehouse.gov/omb/memoranda/m02-10.pdf>]. For OMB's description of PART changes for the FY2005 budget proposal, see U.S. Office of Management and Budget, "Completing the Program Assessment Rating Tool (PART) for the FY2005 Review Process," Budget Procedures Memorandum No. 861, Attachment A, listed on OMB's website at [<http://www.whitehouse.gov/omb/budget/fy2005/part.html>], available at link that follows the text "PART instructions for the 2005 Budget are...", at [<http://www.whitehouse.gov/omb/budget/fy2005/pdf/bpm861.pdf>].

²⁹ For a brief description of this activity, see U.S. Office of Management and Budget, "Completing the Program Assessment Rating Tool (PART) for the FY2005 Review Process," Budget Procedures Memorandum No. 861, p. 4.

³⁰ U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2005, Analytical Perspectives*, pp. 13. NAPA is a congressionally chartered nonprofit corporation. For more information about this type of organizational entity, see CRS Report RL30340, *Congressionally Chartered Nonprofit Organizations ("Title 36 Corporations")*: *What They Are and How Congress Treats Them*, by Ronald C. Moe.

³¹ U.S. Office of Management and Budget, "Program Performance Assessments for the FY 2004 Budget," Memorandum for Heads of Executive Departments and Agencies from Mitchell E. Daniels Jr., p. 2.

the questions OMB asks in advance of making judgments, and opens up any subjectivity in that process for discussion and debate.³²

OMB career staff are not necessarily the only potential sources for subjectivity in completing PART assessments. Subjectivity in completing the PART questionnaire and determining PART scores could potentially also be introduced by White House, OMB, and other political appointees.

Furthermore, in a guidance document for the FY2005 and FY2006 PARTs, OMB has noted that performance measurement in the public sector, and by extension the PART, have limitations, because:

information provided by performance measurement is just part of the information that managers and policy officials need to make decisions. Performance measurement must often be coupled with evaluation data to increase our understanding of why results occur and what value a program adds. Performance information cannot replace data on program costs, political judgments about priorities, creativity about solutions, or common sense. A major purpose of performance measurement is to raise fundamental questions; the measures seldom, by themselves, provide definitive answers.³³

In OMB's guidance for the FY2006 PART, OMB stated that "[t]he PART rel[ies] on objective data to assess programs."³⁴ Former OMB Director Mitchell Daniels Jr. also reportedly stated, with release of the President's FY2004 budget proposal, that "[t]his is the first year in which ... a serious attempt has been made to evaluate, impartially on an ideology-free basis, what works and what doesn't."³⁵ Other points of view regarding how the PART was used are discussed later in this report, in the section titled "Third Party Assessments of the PART."

OMB Use of PART Ratings

In the President's FY2005 budget proposal, OMB stated that PART ratings are intended to "affect" and "inform" budget decisions, but that "PART ratings do not

³² Ibid., p. 3.

³³ This document also identified six "common performance measurement challenges" and "possible strategies for addressing them." To the extent that these challenges make it difficult to assess the results of agency activities, performance measures and the PART might be subject to some validity questions; e.g., whether the chosen measures or PART validly assess the effectiveness of a program. See U.S. Office of Management and Budget, "Performance Measurement Challenges and Strategies," June 18, 2003, available at [http://www.whitehouse.gov/omb/part/challenges_strategies.pdf].

³⁴ Similar language was also included in PART guidance for the President's FY2004 and FY2005 budget proposals. For the FY2006 guidance, see U.S. Office of Management and Budget, "Instructions for the Program Assessment Rating Tool," March 22, 2004, available at [http://www.whitehouse.gov/omb/part/2006_part_guidance.pdf].

³⁵ Quoted in Bridgette Blair, "Investing in Performance: But Being Effective Doesn't Always Pay," *Federal Times*, Feb. 10, 2003.

result in automatic decisions about funding.”³⁶ In OMB’s guidance for the FY2004 PART, for example, OMB said:

FY 2004 decisions will be fundamentally grounded in program performance, but will also continue to be based on a variety of other factors, including policy objectives and priorities of the Administration, and economic and programmatic trends.³⁷

In addition, OMB’s FY2006 PART guidance states that

[t]he PART is a diagnostic tool; the main objective of the PART review is to improve program performance. The PART assessments help link performance to budget decisions and provide a basis for making recommendations to improve results.³⁸

The President’s budget proposals for FY2004 and FY2005 both indicated that the PART process influenced the President’s recommendations to Congress.³⁹

Third-Party Assessments of the PART

Performance Institute

An analysis of the Bush Administration’s FY2005 PART assessments by the Performance Institute, a for-profit corporation that has broadly supported the President’s Management Agenda, stated that “PART scores correlated to funding changes demonstrates an undeniable link between budget and performance in FY

³⁶ U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2005, Analytical Perspectives*, pp. 12-13.

³⁷ U.S. Office of Management and Budget, “Program Performance Assessments for the FY 2004 Budget,” Memorandum for Heads of Executive Departments and Agencies from Mitchell E. Daniels Jr., p. 4.

³⁸ U.S. Office of Management and Budget, “Completing the Program Assessment Rating Tool (PART) for the FY2005 Review Process,” Budget Procedures Memorandum No. 861, p. 1.

³⁹ For media coverage of the PART’s impact on the President’s FY2004 proposals, see Amelia Gruber, “Poor Performance Leads to Budget Cuts at Some Agencies,” *GovExec.com*, Feb. 3, 2003, available at [<http://www.govexec.com/dailyfed/0203/020303a1.htm>]. For media coverage of the PART’s impact on the President’s FY2005 proposals, see Amelia Gruber, “Bush Seeks \$1 Billion in Cuts for Subpar Programs,” *GovExec.com*, Jan. 30, 2004, available at [<http://www.govexec.com/dailyfed/0104/013004a1.htm>]; Christopher Lee, “OMB Draws a Hit List of 13 Programs It Calls Failures,” *Washington Post*, Feb. 11, 2004, p. A29; and Jonathan Weisman, “Deficit is \$521 Billion in Bush Budget,” *Washington Post*, Feb. 2, 2004, p. A01, available at [<http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&contentId=A4093-2004Feb1¬Found=true>]. The latter Web page contains links to two PDF files, reportedly supplied by OMB, that describe the Bush Administration’s proposed cuts and rationales for cuts. The Web addresses for these links are [<http://www.washingtonpost.com/wp-srv/politics/shoulders/budget05cuts.pdf>] and [<http://www.washingtonpost.com/wp-srv/politics/shoulders/budget05cuts2.pdf>].

'05.’⁴⁰ The Performance Institute noted that the President made the following budget proposals for FY2005:

- Programs that OMB judged “Effective” were proposed with average increases of 7.18%;
- “Moderately Effective” programs were proposed with average increases of 8.27%;
- “Adequate” programs were proposed with decreases of 1.64%;
- “Ineffective” programs were proposed with average decreases of 37.68%; and
- “Results Not Demonstrated” programs were proposed with average decreases of 3.69%.

The Performance Institute further asserted that the PART had captured the attention of federal managers, resulted in improved performance management, resulted in better outcome measures for programs, and served as a “quality control” tool for GPRA.⁴¹ The company also asserted that Congress, which had not yet engaged in the PART process, should do so.

Scholarly Assessments

The PART. According to a news report, one prominent scholar in the area of program evaluation offered a mixed assessment of the PART:

Some critics call PART a blunt instrument. But Harry R. Hatry, the director of the public-management program at the Urban Institute, a Washington think tank, said the administration appears to be making a genuine effort to evaluate programs. He serves on an advisory panel for the PART initiative. “All of this is pretty groundbreaking,” he said. Mr. Hatry argues that it’s important to examine outcomes for programs, and that spending decisions ought to be more closely tied to such information. That said, he did caution about how far PART can go. “The term ‘effective’ is probably pushing the word a little bit,” he said. “It’s almost impossible to extract in many of these programs ... the effect of the

⁴⁰ The Performance Institute, “Lessons from the Use of Program Assessment Rating Tool (PART) in the ‘05 Budget: Performance Budgeting Starts Driving Decisions and Paying Dividends,” *White Paper Analysis*, undated (Arlington, VA: [2004]), p. 2. Electronic version available at [<http://www.transparentgovernment.org/tg/fy05budget.htm>].

⁴¹ The Performance Institute, “Lessons from the Use of Program Assessment Rating Tool (PART) in the ‘05 Budget: Performance Budgeting Starts Driving Decisions and Paying Dividends.” The Performance Institute further asserted that “[s]ince June 2003, when a joint Executive-Legislative forum exposed that Congress was barely even aware of the PART, OMB officials have reached out to the Hill and educated staffers on the PART’s value” (p. 4). For media coverage of that forum, see Amelia Gruber, “OMB Ratings Have Little Impact on Hill Budget Decisions,” *GovExec.com*, June 13, 2003, available at [<http://www.govexec.com/dailyfed/0603/061303a1.htm>].

federal expenditures.” Ultimately, while Mr. Hatry is enthusiastic about adding information to the budget-making process, he holds no illusions that this will suddenly transform spending decisions in Washington. “Political purpose,” he said, “is all over the place.”⁴²

Scholars have also begun to analyze the PART using sophisticated statistical techniques, including regression analysis.⁴³ One team investigated “the role of merit and political considerations” in how PART scores might have influenced the President’s budget recommendations to Congress for FY2004 and FY2005 for individual programs.⁴⁴ In summary, they found that PART scores were positively correlated with the President’s recommendations for budget increases and decreases (i.e., a higher PART score was associated with a higher proposed budget increase, after controlling for other variables). The team also found what they believed to be some evidence (i.e., statistically significant regression coefficients) that politics may have influenced the budget recommendations that were made, and how the PART

⁴² Erik W. Robelen, “Itemizing the Budget,” *Education Week*, March 5, 2003, p. 1. Hatry provided similar feedback to OMB in comments on the PART, as a member of the PMAC. See [http://www.whitehouse.gov/omb/budintegration/pmac_030303comments.pdf], “Comments and Suggestions, PART 2004 and 2005,” March 3, 2003, p. 10. Furthermore, *Government Executive* magazine reported that:

Hatry and other evaluation experts see OMB’s effort as a mixed blessing. On the one hand, they are concerned that the initiative is being touted as a way to measure effectiveness, when, in fact, it is largely focused on management issues and lacks the sophisticated analysis needed to truly assess complicated federal programs. On the other hand, they say, if the initiative is successful, it could create a groundswell for more thorough evaluations.

(See Matthew Weinstock, “Under the Microscope,” *Government Executive*, pp. 39-40.) The Urban Institute is a public nonprofit 501(c)(3) corporation.

⁴³ Regression analysis is a statistical technique that attempts to estimate the impact of a change in one “independent” variable on another “dependent” variable, while holding other independent variables constant. For example, a regression model might attempt to estimate how a change in a program’s PART score (independent variable) is associated with changes in the President’s budget recommendations (dependent variable; e.g., percentage increase or decrease in a program’s budget, compared to the previous year).

⁴⁴ For the team’s paper about FY2004 PART assessments, see John B. Gilmour and David E. Lewis, “Does Performance Budgeting Work? An Examination of OMB’s PART Scores,” *Public Administration Review* (forthcoming [2005]), manuscript dated Apr. 19, 2004. Available in electronic PDF or hard copy from this CRS report’s author. This study relied in part on previous work by the same authors: John B. Gilmour and David E. Lewis, “Political Appointees and the Quality of Federal Program Management,” *American Politics Research* (forthcoming [2005]), available at [http://www.wws.princeton.edu/research/papers/10_03_dl.pdf]. A brief summary of the latter paper is available at [http://www.wws.princeton.edu/policybriefs/lewis_appointees.pdf]. A subsequent unpublished paper that analyzed PART assessments for both FY2004 and FY2005 (John B. Gilmour and David E. Lewis, “Assessing Performance Assessment for Budgeting: The Influence of Politics, Performance, and Program Size in FY2005” (unpublished manuscript [2004])) was presented to the 2004 annual meeting of the American Political Science Association in Chicago, IL, Aug. 2004, and is available in hard copy from this CRS report’s author.

was used, for FY2004, but not for FY2005. They also found what they believed to be evidence that PART scores appeared to have influence for “small-sized” programs (less than \$75 million) and “medium-sized” (between \$75 and \$500 million) programs, but not for large programs.⁴⁵

PART and Performance Budgeting. Observers have generally considered the PART to be a form of “performance budgeting,” a term that does not have a standard definition.⁴⁶ In general, however, most definitions of performance budgeting involve the use of performance information and program evaluations during a government’s budget process. Scholars have generally supported the use of performance information in the budget process, but have also noted a lack of consensus on how the information should be used and that performance budgeting has not been a panacea. In state governments, for example:

Practitioners frequently acknowledge that the process of developing measures can be useful from a management and decision-making perspective. Budget officers were asked to indicate how effective the development and use of performance measures has been in effecting certain changes in their state across a range of items, from resource allocation issues, to programmatic changes, to cultural factors such as changing communication patterns among key players.... Many respondents were willing to describe performance measurement as “somewhat effective,” but few were more enthusiastic.... Most markedly, few were willing to attach performance measures to changes in appropriation levels.... Legislative budget officers ranked the use of performance measures especially low in [effecting] cost savings and reducing duplicative services.... Slightly more than half the respondents “strongly agreed” or “agreed” when asked whether the implementation of performance measures had improved communication between agency personnel and the budget office and between agency personnel and legislators.⁴⁷

⁴⁵ The authors attempted to replicate earlier analysis on PART and program size done by GAO, described later in this report.

⁴⁶ *Performance budgeting* has been variously defined by many scholars and practitioners. For an overview, see Philip G. Joyce, “Performance-Based Budgeting,” in Roy T. Meyers, ed., *Handbook of Government Budgeting* (San Francisco: Jossey-Bass, 1999), pp. 597-619. In a widely cited study of performance budgeting at the state level, the term was defined as “a process that requests quantifiable data that provide meaningful information about program outputs and outcomes in the budget process.” (Katherine G. Willoughby and Julia E. Melkers, “Assessing the Impact of Performance Budgeting: A Survey of American States,” *Government Finance Review*, vol. 17, no. 2 (Apr. 2001), pp. 25.) More recently, one authority in the field offered two “polar” definitions: a *broad definition* (“a performance budget is any budget that presents information on what agencies have done or expect to do with the money provided to them”) and a *strict definition* (“a performance budget is only a budget that explicitly links each increment in resources to an increment in outputs or other results”). (Allen Schick, “The Performing State: Reflection on an Idea Whose Time Has Come but Whose Implementation Has Not,” *OECD Journal on Budgeting*, vol. 3, no. 2 (Nov. 2003), p. 101.)

⁴⁷ Katherine G. Willoughby and Julia E. Melkers, “Assessing the Impact of Performance Budgeting: A Survey of American States,” pp. 27, 28, 30.

Another scholar asserted that, among other things, “[p]erformance budgeting is an old idea with a disappointing past and an uncertain future,” and that “it is futile to reform budgeting without first reforming the overall [government] managerial framework.”⁴⁸

Government Accountability Office

GAO recently undertook a study of how OMB used the PART for the FY2004 budget.⁴⁹ Specifically, GAO examined: (1) how the PART changed OMB’s decision-making process in developing the President’s FY2004 budget request; (2) the PART’s relationship to the [Government Performance and Results Act] planning process and reporting requirements; and (3) the PART’s strengths and weaknesses as an evaluation tool, including how OMB ensured that the PART was applied consistently.⁵⁰

GAO asserted that the PART helped to “structure and discipline” how OMB used performance information for program analysis and the executive branch budget development process,⁵¹ made OMB’s use of performance information more transparent, and “stimulated agency interest in budget and performance integration.”⁵² However, GAO noted that “only 18 percent of the [FY2004 PART] recommendations had a direct link to funding matters.”⁵³ GAO also concluded “the more important role of the PART was not in making resource decisions but in its support for recommendations to improve program design, assessment, and management.”⁵⁴

More fundamentally, GAO contended that the PART is “not well integrated with GPRA — the current statutory framework for strategic planning and reporting.” Specifically, GAO said:

OMB has stated its intention to modify GPRA goals and measures with those developed under the PART. As a result, OMB’s judgment about appropriate goals and measures is substituted for GPRA judgments based on a community of stakeholder interests.... Many [agency officials] view PART’s program-by-program focus and the substitution of program measures as detrimental to their

⁴⁸ Allen Schick, “The Performing State: Reflection on an Idea Whose Time Has Come but Whose Implementation Has Not,” p. 100.

⁴⁹ U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB’s Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, GAO-04-174, Jan. 2004.

⁵⁰ *Ibid.*, p. 3.

⁵¹ For an overview of that process, see CRS Report RS20175, *Overview of the Executive Budget Process*, and CRS Report RS20179, *The Role of the President in Budget Development*, both by Bill Heniff Jr.

⁵² U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB’s Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, p. 4.

⁵³ *Ibid.*, p. 12.

⁵⁴ *Ibid.*, p. 11.

GPRA planning and reporting processes. OMB's effort to influence program goals is further evident in recent OMB Circular A-11 guidance that clearly requires each agency to submit a performance budget for fiscal year 2005, which will replace the annual GPRA performance plan.⁵⁵

Notably, GPRA's framework of strategic planning, performance reporting, and stakeholder consultation prominently includes consultation with Congress. Furthermore, GAO said:

Although PART can stimulate discussion on program-specific performance measurement issues, it is not a substitute for GPRA's strategic, longer-term focus on thematic goals, and on department- and governmentwide crosscutting comparisons. Although PART and GPRA serve different needs, a strategy for integrating the two could help strengthen both.⁵⁶

GAO performed regression analysis on the Bush Administration's PART scores and funding recommendations.⁵⁷ In particular, GAO estimated the relationship of overall PART scores on the President's recommended budget changes for FY2004 (measured by percentage change from FY2003) for two separate subsets of the programs that OMB assessed with the PART for FY2004. For *mandatory* programs, GAO found no statistically significant relationship between PART scores and proposed budget changes.⁵⁸ For *discretionary* programs as an overall group, GAO found a statistically significant, positive relationship between PART scores and proposed budget changes.⁵⁹ However, when GAO ran separate regressions on small, medium, and large discretionary programs, GAO found a statistically significant, positive relationship only for small programs.

GAO also came to the following determinations:

⁵⁵ U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, pp. 6-7.

⁵⁶ U.S. General Accounting Office, *Performance Budgeting: OMB's Program Assessment Rating Tool Presents Opportunities and Challenges for Budget and Performance Integration (Highlights, inside front cover)*.

⁵⁷ See Appendix I of the GAO report, pp. 41-46, for a detailed description of GAO's methodology and findings. GAO performed separate regression analyses for mandatory and discretionary programs, as well as for programs of "small," "medium," and "large" size. GAO also analyzed the effect of the PART's component scores on proposed budget changes.

⁵⁸ *Ibid.*, p. 42 (n = 27). *Mandatory* programs, which are distinguished from *discretionary* programs, are associated with federal spending that is governed by substantive legislation, as opposed to annual appropriations acts. Policies and programs involving *discretionary* spending are implemented in the context of annual appropriations acts. For more information, see CRS Report 98-721, *Introduction to the Federal Budget Process*, by Robert Keith and Allen Schick.

⁵⁹ *Ibid.*, p. 43 (n = 196). A one-point increase in overall PART score was associated with a 0.54% increase in proposed budget change (e.g., a 3.00% proposal would increase to 3.54%).

- OMB made sustained efforts to ensure consistency in how programs were assessed for the PART, but OMB staff nevertheless needed to exercise “interpretation and judgment” and were not fully consistent in interpreting the PART questionnaire (pp. 17-19).
- Many PART questions contained subjective terms that contributed to subjective and inconsistent responses to the questionnaire (pp. 20-21).⁶⁰
- Disagreements between OMB and agencies on appropriate performance measures helped lead to the designation of certain programs as “Results Not Demonstrated” (p. 25).⁶¹
- A lack of performance information and program evaluations inhibited assessments of programs (pp. 23-24).
- The way that OMB defined *program* may have been useful for a PART assessment, but “did not necessarily match agency organization or planning elements” and contributed to the lack of performance information (pp. 29-30).⁶²

In response to these issues, GAO recommended that OMB take several actions, including centrally monitoring agency implementation and progress on PART recommendations and reporting such progress in OMB’s budget submission to Congress; continuing to improve the PART guidance; clarifying expectations to agencies on how to allocate scarce evaluation resources; attempting to generate early in the PART process an ongoing, meaningful dialogue with congressional appropriations, authorization, and oversight committees about what OMB considers the most important performance issues and program areas; and articulating and implementing an integrated, complementary relationship between GPRA and the

⁶⁰ For example, GAO’s report stated:

Some agency officials claimed that having multiple statutory goals disadvantaged their programs. Without further guidance, subjective terminology can influence program ratings by permitting OMB staff’s views about a program’s purpose to affect assessments of the program’s design and purpose.

Ibid, p. 20. For more discussion of the PART regarding this issue, and in the context of a health-related “program,” see CRS Report RL32546, *Title VII Health Professions Education and Training: Issues in Reauthorization*, by Sarah A. Lister, Bernice Reyes-Akinbileje, and Sharon Kearney Coleman, under the heading “The Effectiveness of Title VII Programs.”

⁶¹ As GAO noted, during the PART process, OMB created the “Results Not Demonstrated” category when agencies and OMB could not agree on long-term and annual performance measures or if performance information for a program was judged inadequate by OMB.

⁶² GAO noted, for example, that OMB’s aggregation and disaggregation of separate programs into a “PART program” sometimes (1) made it difficult to select measures for “PART programs” that had multiple missions or (2) ignored the context in which programs operate.

PART.⁶³ In OMB’s response, OMB Deputy Director for Management Clay Johnson III stated “We will continually strive to make the PART as credible, objective, and useful as it can be and believe that your recommendations will help us to that. As you know, OMB is already taking actions to address many of them.”⁶⁴

In addition, GAO suggested that while Congress has several opportunities to provide its perspective on performance issues and performance goals (e.g., when establishing or reauthorizing a program, appropriating funds, or exercising oversight), “a more systematic approach could allow Congress to better articulate performance goals and outcomes for key programs of major concern” and “facilitate OMB’s understanding of congressional priorities and concerns and, as a result, increase the usefulness of the PART in budget deliberations.” Specifically, GAO suggested

that Congress consider the need for a strategy that could include (1) establishing a vehicle for communicating performance goals and measures for key congressional priorities and concerns; (2) developing a more structured oversight agenda to permit a more coordinated congressional perspective on crosscutting programs and policies; and (3) using such an agenda to inform its authorization, oversight, and appropriations processes.⁶⁵

Potential Criteria for Evaluating the PART or Other Program Evaluations

Concepts for Evaluating a Program Evaluation

Previous sections of this report discussed how the PART is structured, how it has been used, and how various actors have assessed its design and implementation. This section discusses potential criteria for evaluating the PART or other program evaluations, which might be considered by Congress during the budget process, in oversight of federal agencies and programs, and regarding legislation that relates to program evaluation.⁶⁶ Should Congress focus on the question of criteria, the program evaluation and social science literature suggests that three standards or criteria may be helpful: the concepts of *validity*, *reliability*, and *objectivity*.

⁶³ U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB’s Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, pp. 36-37.

⁶⁴ *Ibid.*, p. 65.

⁶⁵ *Ibid.*, p. 36.

⁶⁶ With regard to legislation, the PART has been prominently mentioned and cited in the context of the proposed Commission on the Accountability and Review of Federal Agencies (CARFA) Act (S. 1668/H.R. 3213, and similar language in budget process reform bills including H.R. 3800/S. 2752 and H.R. 3925), as well as the proposed Program Assessment and Results (PAR) Act (H.R. 3826).

- *Validity* has been defined as “the extent to which any measuring instrument measures what it is intended to measure.”⁶⁷ For example, because the PART is supposed to measure the effectiveness of federal programs, its validity turns on the extent to which PART scores reflect the actual “effectiveness” of those programs.⁶⁸
- *Reliability* has been described as “the relative amount of random inconsistency or unsystematic fluctuation of individual responses on a measure”; that is, the extent to which several attempts at measuring something are consistent (e.g., by several human judges or several uses of the same instrument).⁶⁹ Therefore, the degree to which the PART is reliable can be illustrated by the extent to which separate applications of the instrument to the same program yield the same, or very similar, assessments.
- *Objectivity* has been defined as “whether [an] inquiry is pursued in a way that maximizes the chances that the conclusions reached will be true.”⁷⁰ Definitions of the word also frequently suggest concepts of fairness and absence of bias. The opposite concept is *subjectivity*, suggesting, in turn, concepts of bias, prejudice, or unfairness. Therefore, making a judgment about the *objectivity* of the PART or

⁶⁷ See Edward G. Carmines and James Woods, “Validity,” in Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao, eds., *The SAGE Encyclopedia of Social Science Research Methods*, vol. 3 (Thousand Oaks, CA: SAGE Publications, 2004), p. 1171. The authors elaborate thus: “Thus, the measuring instrument itself is not validated, but the measuring instrument [is validated] in relation to the purpose for which it is being used.”

⁶⁸ For another example, a test may be given to a student that is to assess his or her knowledge of chemistry at a certain level. However, if the test does not assess that knowledge accurately, for that level (e.g., it tests only a narrow aspect of what was supposed to be taught, or turns out to effectively test something else), the test would not necessarily be a *valid* one.

⁶⁹ For example, if a test, on average, accurately assesses the knowledge of a student, the test still may be judged not reliable, viewed either on its own or compared to another test. Suppose a student’s “true” knowledge is equivalent to a 75 score out of 100. Two different styles of tests, if they were applied many times to the same student, might both find, on average, a 75 score. But if one test results in scores that range in a uniform distribution between 73 and 77, while the other test results in scores that result in a uniform distribution between 51 and 100, one could argue that the first test is more reliable as a measure of the student’s knowledge, because it exhibits less fluctuation. If only the second test were realistically available to be used, and if that were the most reliable alternative available, the second test could be judged reliable or unreliable, depending on an observer’s viewpoint and the purpose the test is intended to address. See Peter Y. Chen and Autumn D. Krauss, “Reliability,” in Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao, eds., *The SAGE Encyclopedia of Social Science Research Methods*, vol. 3, p. 952, and Michael Scriven, *Evaluation Thesaurus*, 4th ed. (Newbury Park, CA: SAGE Publications, 1991), p. 309.

⁷⁰ For more information and criticisms of the concept, see Martyn Hammersley, “Objectivity,” in Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao, eds., *The SAGE Encyclopedia of Social Science Research Methods*, vol. 2, pp. 750-751.

its implementation “involves judging a course of inquiry, or an inquirer, against some rational standard of how an inquiry *ought to have been pursued in order to maximize the chances of producing true findings*” (emphasis in original).⁷¹

Although these three criteria can each be considered individually, in application they may prove to be highly interrelated. For example, a measurement tool that is subjectively applied may yield results that, if repeated, are not consistent or do not seem reliable. Conversely, a lack of reliable results may suggest that the instrument being used may not be valid, or that it is not being applied in an objective manner. In these situations, further analysis is typically necessary to determine whether problems exist and what their nature may be.

Evaluating the PART

With regard to the PART, the Administration has made numerous assessments regarding program effectiveness. But how should one *validly, reliably, and objectively* determine a program is *effective*? Should Congress wish to explore these issues regarding the PART or other evaluations, Congress might assess the extent to which the assessments have been, or will be, completed *validly, reliably, and objectively*.

Different observers will likely have different views about the validity, reliability, and objectivity of OMB’s PART instrument, usage, and determinations. Nonetheless, some previous assessments of the PART suggest areas of particular concern. For example, in its study of the PART, GAO reported that one of the two reasons why programs were designated by the Administration as “results not demonstrated” (nearly 50% of the 234 programs assessed for FY2004) was that OMB and agencies disagreed on how to assess agency program performance, as represented by “long-term and annual performance measures.”⁷² Different officials in the executive branch appeared to have different conceptions of what the appropriate goals of programs, and measures to assess programs, should be — raising questions about the *validity* of the instrument. It is reasonable to conclude that actors outside the executive branch, including Members of Congress, citizens, and interest groups, may have different perspectives and judgments on appropriate program goals and measures. Under GPRA, stakeholder views such as these are required to be solicited by statute. Under the PART, however, the role and process for stakeholder participation appears less certain.

Other issues that GAO identified could be interpreted as relating to the PART instrument’s *validity* in assessing program effectiveness (e.g., OMB definitions of specific programs inconsistent with agency organization and planning); its *reliability*

⁷¹ Ibid., p. 750. Thus, analysts often ask whether a given instrument can be improved (i.e., whether the instrument’s chances of reaching valid and reliable findings have been maximized). An implication of these terms is that it is possible for an instrument to be objective, but not a valid measure of what it is intended to measure.

⁷² See U.S. General Accounting Office, *Performance Budgeting: Observations on the Use of OMB’s Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, p. 25.

in making consistent assessments and determinations (e.g., inconsistent application of the instrument across multiple programs); and its *objectivity* in design and usage. To illustrate with some potential examples of objectivity issues, subjectivity could arguably be resident in a number of PART questions, including, among others, when OMB conducted its assessment for FY2005.⁷³

- whether a program is “excessively” or “unnecessarily” ... “redundant or duplicative of any other Federal, State, local, or private effort” [question 1.3, p. 22];⁷⁴
- whether a program’s design is free of “major flaws” [question 1.4, p. 23];⁷⁵
- whether a program’s performance measures “meaningfully” reflect the program’s purpose [question 2.1, p. 25];⁷⁶ and
- whether a program has demonstrated “adequate” progress in achieving long-term performance goals [question 4.1, p. 47].

Use of such terms that, in the absence of clear definitions, are subject to a variety of interpretations can raise questions about the objectivity of the instrument and its ratings.

In one of its earliest publications on the PART, OMB said that “[w]hile subjectivity can be minimized, it can never be completely eliminated regardless of the method or tool.⁷⁷ OMB went on to say, though, that the PART “makes public and transparent the questions OMB asks in advance of making judgments, and opens up any subjectivity in that process for discussion and debate.” That said, the PART and its implementation to date nevertheless appear to place much of the process for debating and determining program goals and measures squarely within the executive branch.

⁷³ For OMB’s FY2005 PART questions and guidance, see U.S. Office of Management and Budget, “Completing the Program Assessment Rating Tool (PART) for the FY2005 Review Process,” Budget Procedures Memorandum No. 861 from Richard P. Emery Jr., May 5, 2003, listed on OMB’s website at [<http://www.whitehouse.gov/omb/budget/fy2005/part.html>], available at link that follows the text “PART instructions for the 2005 Budget are...” at [<http://www.whitehouse.gov/omb/budget/fy2005/pdf/bpm861.pdf>]. Page references in the bulleted text refer to this memorandum.

⁷⁴ OMB’s FY2006 PART guidance contains the same language. See U.S. Office of Management and Budget, *Instructions for the Program Assessment Rating Tool* (undated), p. 15, available at [http://www.whitehouse.gov/omb/part/2006_part_guidance.pdf].

⁷⁵ OMB’s FY2006 PART guidance contains the same language. See *ibid.*, p. 16.

⁷⁶ See *ibid.*, p. 18. OMB’s website defines *meaningful* as “measur[ing] the right thing — usually the outcome the program is intended to achieve.” See “PART Frequently Asked Question” #14, available at [http://www.whitehouse.gov/omb/part/2004_faq.html].

⁷⁷ U.S. Office of Management and Budget, “Program Performance Assessments for the FY 2004 Budget,” p. 2.