ANALYZING THE FEDERAL GOVERNMENT’S PROGRAM ASSESSMENT RATING TOOL (PART) IN DETERMINING BUDGET ALLOCATIONS

A Thesis
Submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy
in the Georgetown Public Policy Institute

By

Michael B. Hilburg, B.S.B.A.

Washington, D.C.
April 15, 2010
ANALYZING THE FEDERAL GOVERNMENT’S PROGRAM ASSESSMENT RATING TOOL (PART) IN DETERMINING BUDGET ALLOCATIONS

Michael B. Hilburg, B.S.B.A.

Thesis Advisor: Joydeep Roy, Ph.D.

ABSTRACT

This thesis will examine the effect of Office of Management and Budget (OMB) Program Assessment Rating Tool (PART) scores on changes in budget allocation in the President’s Budget during George W. Bush’s Presidency. Numerous OMB press releases indicate White House usage of PART scores as a budget allocation tool; however, there has been no study to examine this relationship over the duration of President Bush’s term. Utilizing publically available information from OMB, I will observe and analyze the relationship by employing two OLS regression models.

Initial, simple regression models showed PART scores were positively correlated with subsequent year Presidential budget recommendations. However, complex models that account for variances during different years illustrate no consistent use of PART scores in the budgeting process.
**Table of Contents**

Chapter 1: Introduction ................................................................................................... 4  
  Background: Office of Management and Budget................................................. 6  
  Background: Program Assessment Rating Tool (PART) ................................... 7  
  The Current Debate about PART Effectiveness .............................................. 9  

Chapter 2: Literature Review: ....................................................................................... 13  
  Gilmour & Lewis – 2006(a): .......................................................... 13  
  Gilmour & Lewis – 2006(b): ....................................................................... 14  
  Tat-Kei Ho – 2005: .............................................................................. 15  
  Posner & Fantone – 2007: ................................................................. 17  

Chapter 3: Research Design .......................................................................................... 19  
  Hypothesis ............................................................................................ 19  
  Data Source .......................................................................................... 19  
  Variables.................................................................................................. 20  
    Primary Variables of Interest.............................................................. 20  
    Control Variables ............................................................................. 23  
  Models .................................................................................................. 29  
  Descriptive Statistics ........................................................................... 29  

Chapter 4: Results and Discussion ................................................................................ 33  
  Model 1 .................................................................................................. 33  
    Initial Results .................................................................................. 34  
    Accounting for Time ....................................................................... 37  
  Model 2 .................................................................................................. 47  

Chapter 5: Conclusion and Policy Implications ............................................................ 51  
  Limitations and Direction for Future Research .............................................. 53  

Endnotes:........................................................................................................... 55  

Bibliography and Referenced Documents: ............................................................. 57
Listing of Tables and Charts

Table 1: Number of Programs Receiving Initial or Subsequent Reviews .......................9
Table 2: Program Ratings by Year (Initial Reviews Only) .............................................30
Table 3: Change in Magnitude Between Reviews .........................................................31
Chart 1: Levels of Rating Improvement for Re-Reviewed Programs ............................32
Table 4: Initial Regression Results ................................................................................34
Chart 2: Average Allocation Increase by Ratings ..........................................................35
Table 5: Abbreviated Results: PART Scores Separated by Political Bias .....................37
Chart 3: Ratings Given per Year – Initial Reviews Only ..............................................38
Table 6: Regression Results, Accounting for Year of Review ....................................39
Table 7: Regression Results, Interactive Model ............................................................41
Table 8: Interactive Coefficients ....................................................................................43
Chart 4: Interactive Model – Percentage Budget Allocation Change by Program Rating ....44
Chart 5: Interactive Model – Percentage Budget Allocation Change by Year ..............46
Table 9: Regression Model Results, Model 2 ...............................................................48
Chapter 1: Introduction

Unlike business enterprises, Government programs generally do not earn profits, making it difficult to measure effectiveness. Government and businesses frequently hire consultants to analyze processes, benchmark operations versus comparable organizations, and create internal performance management metrics to judge efficiency and results. To be truly successful in ensuring effective resource management, these mechanisms need to consider criteria directly relevant to decision makers and provide a method to improve the status quo.

To achieve these goals and promote efficient delivery of services, the government over the years has implemented a variety of performance management tactics. Prior and current Federal Government management programs include:

- President's Committee on Administrative Management (the Brownlow Commission) – 1937
- Planning, Programming and Budgeting System – introduced by President Lyndon Johnson – used widely by the Department of Defense
- Management by Objective – introduced by President Nixon
- Zero-Based Budgeting – 1977 – introduced by President Carter
- Private Sector Survey on Cost Control (Grace Commission) – 1982-1984 –
undertaken by President Reagan’s Administration

- Government Performance and Review Act (GPRA) – 1993-Current – the first performance measurement program codified into law by Congressional Act
- President’s Management Agenda (PMA) – 2001-Current – introduced by President George W. Bush – includes the Program Assessment and Rating Tool (PART)

In this paper I will look at the most recent wide-scale government performance management program, the Program Assessment Rating Tool (PART). Utilizing publically available information from the White House and the Office of Management and Budget (OMB), I will conduct several OLS regressions to measure the relationship between PART ratings and changes in Presidential budget allocations. The strength and direction of the relationship in question will illustrate the level at which the Executive Branch uses the performance management tool to allocate funds. Based on the results of prior research, I hypothesize a positive relationship between PART ratings and budget allocations (higher rated programs should generally receive larger budget increases). Additionally, I anticipate that the relationship will be small to moderately sized, as the PART is simply one tool in a larger box of decision-making criteria and other factors, such political agenda and bias, likely weigh more heavily in budgeting decisions.

In the past year and a half, the PART has become a topic of much debate as the
Obama Administration considers the tool’s role. Accompanying his FY2010 budget, President Obama laid out his management agenda, “Building a High-Performing Government.” The President’s first initiative in this agenda is to replace the PART with a new performance improvement and analysis framework. While the Administration has yet to announce specifics regarding its new performance management tool, some subject matter experts believe the Administration should not eliminate the PART, but “build on the data collection and reporting within PART… and use that data in a different way.”\(^1\) Additionally, a report by Accenture's Institute for Public Service Value, the Georgetown Public Policy Institute and OMB Watch suggests that PART “be revised and updated to address problems evident in both its design and its implementation.”\(^2\)

Through this paper, I intend to examine the PART’s level of use, or lack thereof, in determining budget allocations. This information can help the Obama Administration understand the PART’s prior usage and determine the optimal level of change necessary to improve the government’s performance management measurements.

**Background: Office of Management and Budget**

The Office of Management and Budget (OMB) is the section of the Executive Office of the President charged with the mission of assisting the President to oversee the preparation of the federal budget. As part of this mission, OMB supervises the
administration of these funds across all Executive Branch agencies. As such, OMB evaluates the effectiveness of agency programs, policies and procedures; sets funding priorities; ensures that agency reports and proposed legislation are consistent with the President’s agenda; and coordinates all aspects of financial management. Given the scope, mission and connection to the President’s mission, OMB is properly positioned to work with the President to integrate a performance management system into budget allocation decisions.

**Background: Program Assessment Rating Tool (PART)**

Following the corporate scandals of the early 21\textsuperscript{th} Century, President Bush sought to increase government accountability (alongside efforts to increase private sector accountability). In 2002, the White House and the Office of Management and Budget (OMB) began a pilot program to rate the management and performance of programs. As of FY2009, OMB has assessed all 1,015 Executive Branch federal programs.

PART was designed under the Performance Improvement Initiative of the PMA with the stated intent of building on the Government Performance and Results Act to enhance program management, encourage continuous improvement, and link performance and budget results. Programs under review are asked a series of standardized of 25 questions, which are weighted under four different sections: 1) Purpose and design (20\% of total score); 2) Planning (10\%); 3) Management (20\%);
and 4) Results (such as whether a program is meeting its long-term and annual goals; 50%). Scores from these categories are totaled and programs are given ratings based on the net sum:

- Effective (100-85% of total possible score),
- Moderately Effective (84-70%),
- Adequate (69-50%),
- Ineffective (49-0%), or
- Results Not Demonstrated – Given if OMB decides, independent of a program’s score, that performance information, performance measures, or both are insufficient or inadequate (e.g., the program has not developed goals or has not collected sufficient data to measure)

Assessments focus on identifying the strengths and weaknesses of individual programs based on program-level data. Although the 25 questions OMB asks are standardized, there is a recognized diversity of federal government functions which need to be measured on slightly different criteria. As such, questions are partially tailored based on the type of program; however, questions remain standard for each program type. The seven program types are Direct Federal, Competitive Grant, Block/Formula Grant, Regulatory, Capital Assets and Service Acquisition, Credit, and Research and Development. In an effort to aid consistency in reporting and analysis, the assessments include: the 25 questions phrased for a Yes-No answer, an assessment
of the program’s on-going improvement plans, comments on completed improvement plans and evidence-based performance measures.

Since the initial assessment, completed in calendar year 2002, OMB has rated all classified 1,105 programs in the federal government at least once, as illustrated in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Program Assessed</td>
<td>196</td>
<td>152</td>
<td>200</td>
<td>210</td>
<td>191</td>
<td>47</td>
<td>19</td>
</tr>
<tr>
<td>Re-reviews</td>
<td>0</td>
<td>80</td>
<td>49</td>
<td>29</td>
<td>36</td>
<td>70</td>
<td>48</td>
</tr>
<tr>
<td>Net Programs Assessed</td>
<td>196</td>
<td>232</td>
<td>249</td>
<td>239</td>
<td>227</td>
<td>117</td>
<td>67</td>
</tr>
</tbody>
</table>

Overtime, some managers seen benefit to the PART and undergone a re-review process with the intent of raising their rating level.4

**The Current Debate about PART Effectiveness**

“Taken seriously, rigorous performance assessment will boost the quality of federal programs, and taxpayers will see more of the results they were promised. What works is what matters, and achievement should determine which programs survive, and which do not.”

– President’s FY2004 Budget, introducing the first use of PART in the budgeting process5

Political commitment to an issue is often measured by the amount spent on a
project; however, spending more does not necessarily improve a program. In modern
governing, results matter – not just spending levels. The PART was created to bridge
this gap between spending and results by offering guidance and detailed levels of
budget analysis.

However, for PART to be effective as a management tool, evidence of its use,
at least by the White House, must be apparent. In February 2003, Carl DeMaio,
Founder of the Performance Institute and former-Presidential advisor, stated that “on
average, the President's 2004 budget proposal rewarded programs deemed ‘effective’
with a 6 percent funding increase, and held those ‘not showing results’ to less than a 1
percent increase.” As indicated by Mr. DeMaio’s announcement, PART marked the
first wide-scale use of a formal link between federal spending and program
performance. Over the next several years, the Bush Administration touted the benefits
of PART as a vehicle for improving program performance. Building on what the
White House must have deemed as a successful and useful tool, in 2007 President
signed Executive Order 13450, which sought to make aspects of the PART permanent.

PART’s success as a useful management tool extends beyond the White House
and into agencies and programs. While certainly not all agency and program managers
found the tool useful, many did. Programs, such as the US Department of the
Treasury’s Office of Technical Assistance or the Department of Housing and Urban
Affairs’ HOME (affordable housing block grant program), were able to change their
respective cultures, measure and track performance outcomes, and deliver enhanced services and information to involved parties.\textsuperscript{8}

While evidence exists illustrating the importance of PART within the White House and program areas, the empirical link between PART results and Congressional budget allocations are very limited. While some of the limitations on Congressional usage are due to structural issues of the budget process, other factors appear to be more relevant and useful for Congress in its budget allocation decisions.

Structurally, the information gathered as part of the PART process remains non-public information determined in discussion between OMB and the respective programs. When information is eventually shared between external stakeholders, it can be incomplete, non-specific or in a different format, making its use difficult. For instance, “it is not that Congress does not understand PART or how it works–it is that [Congress] is not a part of the conversation about performance data and results and is therefore highly skeptical about the reported program results.”\textsuperscript{2} Therefore, it is difficult for Congressional members and staffers to become engaged in and excited about the Executive Branch PART process.

Furthermore, the dynamics of the budget process make it difficult for the PART to have an effect on the final budget. The President’s budget serves as a framework for Congress’ final budget proposal. While the President and OMB have full access to the information contained within the PART assessments, Congress does not. As a result,
when the President’s PART informed budget is forwarded to Congress, the President’s allocation recommendations may be determined to be “dead-on-arrival,” as they were for FY2009, and Congress may seek a different direction with expenditures. The end product may likely be a budget that has been modified to such an extent that the PART informed decisions are no longer included.

Alternatively, given structural issues, lawmakers continue to state their general preference against PART in budget allocation decisions. One senior Appropriations staffer stated, “Lawmakers make funding decisions based on traditional budget justification documents, and pay little attention to the Office of Management and Budget’s recent [PART] evaluations.”

Given the strong empirical link between PART results and the OMB decision-making process, and the tenuous link with the Congressional budget allocation process, I will focus my efforts on analyzing PART’s overall effectiveness as a White House budgeting tool. In specific, I will try to identify the potential statistical relationship between program-level PART scores and allocations in the President’s budget throughout President George W. Bush’s tenure. Additionally, I will test to see if there is a statistically significant relationship between changes in program-level PART scores (due to re-review) and subsequent budget levels.
Chapter 2: Literature Review:

Gilmour & Lewis – 2006(a):

The primary relevant statistical research in the area of the PART and performance budgeting was published in 2006 by John Gilmour (College of William and Mary) and David Lewis (Princeton University). In their study, the authors analyzed the programs in the PART’s inaugural class, to determine an effect of the ratings on allocations in the President’s FY2004 budget.

In nearly all of their models, Gilmour and Lewis noted a program’s PART score had a statistically significant effect on OMB’s budget allocations. When accounting for the political content and age (a proxy for stability) of the program, the authors found PART scores were positively correlated (and statistically significant) with the proposed budgets in traditionally Democratic departments (Department of Commerce, Education, Energy, Housing and Urban Development, Labor, Health and Human Services and the Environmental Protection Agency). From these results, Gilmour and Lewis conclude that while President Bush was likely to favor traditional Republican departments with additional funding, funding increases or reductions for Democratic programs were allocated based on effectiveness as determined by the PART scores.

Gilmour and Lewis quantified the effects of political bias by assessing programs based on the program’s Departmental status (whether the department’s
mission is typically considered Republican or Democratic), and the party affiliation of the President and Congress and if the Executive and Legislative branches were unified under one party at the time of program creation.

Gilmour and Lewis’ study provides excellent analysis of the effectiveness of performance budgeting through the PART during the tool’s first full year. I aim to extend their analysis over the entire length of President Bush’s tenure to see if the PART remained as influential, or if its importance dwindled overtime as other pressing issues required more of the President’s time.

**Gilmour & Lewis – 2006(b):**

In 2006 Gilmour and Lewis released a second study analyzing the effectiveness of the PART as a budget allocation tool. Building upon their previous study, the authors analyzed how changes in PART scores for programs that were assessed in initial year and reassessed the next year reflected changes in budget allocations. Through their research, Gilmour and Lewis found that on average a ten percentage point increase in PART scores led to a four to five percent budget increase for re-reviewed programs. Additionally, the authors found that PART seemed to have a bigger impact on small and medium sized programs, than on larger programs. The supposition behind these results is that larger programs have larger constituencies and political backing. As such, changes in PART ratings would need to overcome a greater inertia to affect budget allocations. Whereas, small and medium programs do not have
the same external support and therefore are more likely to be evaluated based on the 
information White House decision makers (PART data).

As in their first study, Gilmour and Lewis still struggle to control for the effects 
politics has on budget allocations, since politics pervades every aspect of budgeting. 
However, by using changes in PART scores over a 1-year time span as the primary 
independent variable, the authors are able to mitigate some of the effects of politics, as 
the effects of politics on a program-level basis did not likely change too much over that 
one year span.

Gilmour and Lewis’ second study builds off their initial research in a strong 
fashion, contributing additional insight and verifying the results from their initial study. 
I will build upon this research by including a control for program size into both of my 
models and extending the timeline to encompass all reassessments during the Bush 
Administration to determine if the results they found carried forward.

**Tat-Kei Ho – 2005:**

Related to the topic and relevancy of performance measurement and budgeting, 
Alfred Tat-Kei Ho researched the effects of performance management tools in the 
decision making process at the local level. Tat-Kei Ho surveyed mayors of 
Midwestern cities to explore how political concerns and the implementation strategies 
of performance measurement influenced their perception of the tools.

Tat-Kei Ho found that performance measurements were widely adopted in
small and medium-sized cities and that 70% of respondents agreed performance measurements were useful to their City Council in enabling it to set priorities in resource allocation.

To explain why some cities and mayors do not use performance measurements, Tat-Kei Ho identifies two broad, but key, possible ideas, including

1. Possible politicalization – either by their opposition (campaign opponent or resident issue opposition) or by the media (if have negative results); and

2. Structural changes necessary to make performance management effective – performance programs are the most effective when accompanied by strategic planning, goal setting, and regular program review. Additionally, obtaining buy-in from residents can be effort intensive.

The author concludes that the lack of stakeholder involvement in the creation process of performance management criteria may limit the potential success and usage, because citizen involvement and leadership by elected executives can increase the political relevance and importance.

The implications from Tat-Kei Ho’s study are important and should be considered in any potential remodel or any performance management program.

1. At least at the local level, performance management measurements are taken serious by elected officials in their decision making process. While this does not predicate a direct link to their importance at the federal level, it does indicate a
possible relevance to the legislative branch (City Councils at the local level and Congress on a national scale). It is important that for performance management tools to be truly effective, decision makers, whether Congress or program managers, to be vested in the results of the process.

2. Involving additional stakeholders in the process and incorporating in strategic planning, goal setting, and regular program review, can enhance the value of the process. While PART was created by the Executive Branch with little Congressional input, key components of the tool’s score engender goals setting, strategy evaluation and effectiveness reviews, which make the measurements relevant.

**Posner & Fantone – 2007:**

While not statistical in their review of the PART, Paul Posner (George Mason University) and Denise Fantone (U.S. Government Accountability Office) consider the effectiveness of the performance measurement process, based on empirical considerations. The authors argue that PART’s predecessor/contemporary performance evaluation system, the Government Performance and Review Act (GPRA), struggled in its effectiveness, despite its Congressional mandate, because budgeting is an inherently political process that reflects a wide-range of competing needs and interests. Posner and Fantone add that PART enhances the information available to decision makers beyond GPRA, adding value to the process. However,
they argue, “it is unlikely that the broad range of actors whose input is critical to decisions will use performance information unless they believe it is credible and reliable and reflects a consensus about performance goals among a community of interested parties.” (p 12) The authors are contending that to aid the effectiveness of PART, OMB should involve Congressional stakeholders in the creation and implementation process. While involving Congress in the creation of the PART may create buy-in for the program, Posner & Fantone believe that unless Congress is also consulted on the budget formulation process, the tool’s usefulness will not be maximized. The authors warn, however, that even if all relevant stakeholders are included in the process, result expectations should be tempered, due to the subjective nature of allocating limited resources.
Chapter 3: Research Design

Hypothesis

This thesis hypothesizes that programs receiving better PART review scores will receive higher recommended increases in the subsequent Presidential Budget. While I expect this hypothesis to hold over the entire tenure of President Bush’s term in office, I anticipate the benefits received between performing programs (those receiving scores of Effective, Moderately Effective and Adequate) to decrease overtime as the overall number of firms receiving these grades increases. I view this situation in similarity to grade inflation in schools, where the more students (or programs) receive higher grades, the less valuable these measures become in differentiating performance. Likewise, as the value of performance becomes more homogeneous and less valuable, the rarity of non-performing results will become more costly and have strong negative ramifications.

Additionally, this thesis will test a second model to evaluate if programs undergoing the re-reviews process receive a similar hypothesized relative budget increase.

Data Source

The data used in my analysis is public information available from the OMB’s Office of Performance and Personnel Management. Prior regression analysis on the topic is limited to early analysis on the first two years; therefore, it was necessary to
compile several years worth of information into one complete dataset. Published information containing PART scores and ratings by program are largely available online from at OMB’s website or the PART website.\textsuperscript{13} PART’s Annual Updates files are compiled and published every January as part of OMB’s budget analysis.

**Variables**

In order to examine this relationship, a theoretical model must be introduced with various independent variables that will control for additional factors that may affect changes in budget allocations. These variables are included so that cross-year PART score comparisons can be appropriately made. A detailed description of all variables is contained below:

**Primary Variables of Interest**

To test the initial relationship of interests, my primary independent variable is the program-level PART score for the programs reviewed by OMB during calendar years 2003 through 2008. To appropriately measure and differentiate these results I will break down the initial ordinal rankings given by OMB (Effective, Moderately Effective, Adequate, Ineffective and Results Not Demonstrated) into five separate indicator variables. All programs received only one rating during their initial review; therefore, all programs will have only one positive response among the five indicators. These variables will be:
• **Effective**, with a value of 1 representing that a program received an Effective rating; a value of 0 indicates that the program received a different rating (Effective will be the omitted category in Model 1);

• **Moderate**, with a value of 1 representing that a program received an Moderately Effective rating; a value of 0 indicates that the program received a different rating;

• **Adequate**, with a value of 1 representing that a program received an Adequate rating; a value of 0 indicates that the program received a different rating;

• **Ineffective**, with a value of 1 representing that a program received an Ineffective rating; a value of 0 indicates that the program received a different rating;

• **RND**, with a value of 1 representing that a program received a Results Not Demonstrated rating; a value of 0 indicates that the program received a different rating;

To test my secondary relationship, I will analyze the 262 programs that underwent at least one PART re-reviewed to determine if changes in their PART scores have an effect on budget allocations (a net of 312 re-reviews were conducted). For this model, my primary independent variable will be the magnitude of the change in scores. As such, the variable ‘Magnitude’ will have values ranging between -4 and 4. For each reassessed program, I assigned the value of 5 to Effective ratings; 4 to Moderately Effective rating; 3 to Adequate; 2 Ineffective; and 1 to Results Not Demonstrated; and then subtracted the initial scores from the re-review scores. For
example, a program could earn a score of -4 if in its initial rating it was judged to be Effective, but in its subsequent review was rated as Results Not Demonstrated. Likewise, a program that received a PART score of Adequate in its initial rating and Moderately Effective in its re-review would have a Magnitude score of 1 (as 4–3=1).

It is important to note that since program managers generally initiate PART re-evaluations, a program may have significantly strengthened its efficiency and effectiveness on their own initiative, regardless of their PART scores. In this instance, outside knowledge of these improvements by the White House may result in additional funding levels. My research and analysis assumes that the White House’s knowledge of changes in program efficiency and effectiveness is relayed solely through the PART. This assumption can create an omitted variable bias and cause a spurious correlation between scores and budget changes. However, since PART scores serve as a proxy measure of efficiency and effectiveness, improvement created as a reaction to an initial PART score could be seen as a result of the program and thus non-biased.

The primary dependent variable in both models, percentage change in allocated budget levels (referred to as Pchange in the dataset), is derived from the program-level budgets proposed by the President’s budget. Given the diverse size of federal programs, judging the relationship based on net appropriated increases would skew the results toward the larger programs, which all else equal would be expected to receive larger net dollar increases. Therefore, a comparison based on percentage change
provides a less biased measure. Additionally, the use of comparison data between the prior and current fiscal years provides a fixed-effect style comparison, which should remove some omitted factors from the model.

**Control Variables**

While President Bush created the PART to bridge the chasm between performance and budget allocation; however, several other factors could bias the results of my study if not included.

**Model 1 Control Variables**

- **Year of PART Review**: The year a program was reviewed can have an effect on the results of its PART score. To best account for the year of review, I will separate the ordinal year variable into a series of variables indicator variables by calendar year (CY2003, CY2004, CY2005, CY2006, CY2007, CY2008). For each of these indicators, a value of ‘1’ denotes that program was underwent PART review during that year, while a value of ‘0’ indicates the review happened during another year. For both models, CY2008 will serve as the omitted category.

  Reform initiatives, whether in the public or private sector, were created to solve a problem; however, as time progresses and goals changes prior important initiatives fade in relevance. When OMB created the PART, President Bush’s policy agenda was drastically different then it was during the later years of his
Administration. During the early years of the PMA and PART, the country suffered numerous scandals (e.g., Enron, Worldcom, and Tyco) and there were calls for greater accountability. However, during President Bush’s second term the national focus was targeted more towards the wars in Afghanistan and Iraq, as well as the recession. As a result, I would expect the relationship between PART scores and budget allocations to lessen throughout President Bush’s tenure as the political focus shifted away from accountability.

- **Type of Federal Program**: OMB differentiates the questions it asks program managers based on the type of program. Within the PART there are seven programs types: Direct Federal, Competitive Grant, Block/Formula Grant, Regulatory, Capital Assets and Service Acquisition, Credit, and Research and Development. Each category of programs has different output and outcome measurements, and therefore requires different questions and performance measures to determine success. Including program type should enhance the overall significance of the primary relationship by helping to explain portions of the variation.

To account for program type, I have created a series of seven indicator variables with each variable representing one of the program types. For these indicators, a value of ‘1’ represents that the variable is that specified type of program, while a value of ‘0’ indicates that the program is of a different type. Programs, as rated by
OMB, can be rated as multiple types. For example, OMB determined that the Federal Communication Commission’s Universal Service Fund E-Rate program is both a Block/Formula Granting and a Regulatory program because of its broad responsibilities. As such, this program will have ‘1’ as the value in both of these type classifications.

- **Political Bias**: Measuring and controlling for political bias in my analysis is a difficult problem. Because politics pervades the budget process so deeply, I will likely never be able to account fully for its effects. However, in an attempt to account for this bias, I will adopt a technique used by Gilmour and Lewis in their 2006(a) study. In this study, the authors loosely grouped programs based on the Republican Party’s stereotypical disfavoring of the program’s parent department or agency. Gilmour and Lewis grouped the Departments of Commerce, Education, Energy, Health and Human Services, Housing and Urban Development, and Labor, as well as the Environmental Protection Agency, as departments central to the agenda of the Democratic Party, but not the Republican Party. Although exceptions exist in this broad generalization, to avoid an ad hoc approach and without in-depth knowledge on how specific programs may have been affected, I believe it better to match Gilmour and Lewis’ prior work and rely on these traditional party positions. For this indicator variable, programs with a value of ‘1’ are within traditionally Democratic agencies, while a value of ‘0’
pertains to programs within all other agencies.

- **Program Size**: Large federal programs typically have a large constituency, their added resources likely affect sizable number of people making it more difficult to drastically downsize these programs. Conversely, smaller programs are more likely to have wild swings in funding levels; it is easier to eliminate smaller, less established programs due to their lesser level of support. As a result, PART scores may have a more radical percentage effect on budgeted levels for smaller, less established programs that could bias the results of the regression analysis. I would expect that adding program size to the model would enhance the overall significance of the relationship of concern by explaining some of the variation between programs, while controlling for potential outliers.

To measure program size I will look at the size of a program’s funding during the fiscal year before its review. A plotting of the size of program funding shows a very long tail, with several programs receiving hundreds of billions of dollars (e.g., Medicare and Medicaid). To account for these outliers, I will use the natural log of these funding levels (in millions of dollars).

- **Parent Agency Size**: PART reviewed programs fall into 56 parent agencies, 34 of which contain ten or fewer PART reviewed programs. Programs in larger agencies may be small compared to its parent organization (such as the Energy Conservation Investment program within the Department of Defense) and
therefore may not contribute to the department’s profile, but may ultimately obtain a larger budget because of net departmental increases. Whereas, in smaller agencies, individual programs will likely contribute more to the overall percentage and profile of the parent organization, making it less likely that a respective program will see an increase as a free-rider program. Controlling for swings in larger agency and department funding levels should provide greater confidence in identifying funding changes specifically due to PART reviews. To measure this effect I will create an indicator variable to represent the larger parent agencies (value of ‘1’ for larger agencies and a value of ‘0’ for smaller).

**Model 2 Control Variables**

In model 2, I will utilize many of the same control variables, including:

- Program Type;
- Political Bias;
- Program Size; and
- Parent Agency Size

In addition, I created several new control variables to account for the differences between the two models, including:

- **Span**: Including the span of time, in years, between the initial and follow-on PART reviews in the model will allow my analysis to control for the effects timing has on the review process. I believe that the relationship between PART
scores and budget allocations will weaken over time, as the President’s focus shifts from accountability to the nation’s two wars. Likewise, I would hypothesize that managers who take the initiative to quickly make changes within their programs can more easily emphasize to budgetary decision makers the positive effects of their enhancements. Therefore, if a program goes numerous years before enacting OMB’s recommended changes, the overall improvements will be muddied by changes external to their programs. To generate values for this variable, I subtracted the numeric value of the original PART review (e.g., 2003) from the year of the re-review (e.g., 2008). The resulting number (5 in this example) is the variable’s value for that program. The inclusion of the span variable in this model should allow for a more significant relationship between to the primary variables.

• **Previous Rating** *(PrevModerate, PrevAdequate, PrevIneffective and PrevRND)*: Just as a program’s re-reviewed PART score may affect budget allocation increases, the program's initial review results may also have an effect. While the Magnitude variable will measure the overall change in results, these indicator variables provide an indication of the program’s initial results. Programs that were previously Ineffective, but rose one spot in magnitude to Adequate, may earn a higher increase than programs that rose from Moderately Effective to Effective.
It is this relationship these previous rating variable intend to capture. For the second model, PrevEffective serves as the omitted group in the models.

**Models**

The regression models used for my thesis will be multivariate OLS regressions. Utilizing the primary variables of interest and the controls, the initial models appear as follows:

**Model 1:**

\[
\text{Percentage Change in Budget Amount} = \beta_0 + \beta(\text{Rating indicator variables}) + \beta(\text{Program Type}) + \beta(\text{Political Bias indicator Variable}) + \beta(\text{Ln(prior budget amount)}) + \beta(\text{Agency size indicator variable})
\]

**Model 2:**

\[
\text{Percentage Change in Budget Amount} = \beta_0 + \beta(\text{Magnitude}) + \beta(\text{Program Type}) + \beta(\text{Political Bias indicator Variable}) + \beta(\text{Ln(prior budget amount)}) + \beta(\text{Agency size indicator variable}) + \beta(\text{Span between reviews})
\]

**Descriptive Statistics**

The compiled dataset consists of 1,327 data points (PART scores) from 1,015 programs, as delineated by OMB. 223 programs were re-reviewed once; 49 programs were reviewed three times (one initial plus two re-reviews); and one program, the Department of Defense’s Missile Defense program, was reviewed four times. Programs exist within 56 parent departments or agencies, of which 24 of these
departments or agencies contain over 10 programs.

Drilling further into the breakdown of program reviews, as more programs were reviewed under PART, a greater percent of programs received performing ratings (Effective, Moderately Effective, or Adequate). The top of Table 2 shows the number of programs that receive a specific rating level by year, while the bottom half of the table shows the cumulative break down of review results per year by rating level (initial reviews only).

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program Ratings by Year (Initial Reviews Only)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Effective</td>
</tr>
<tr>
<td>Moderately Effective</td>
</tr>
<tr>
<td>Adequate</td>
</tr>
<tr>
<td>Ineffective</td>
</tr>
<tr>
<td>Results Not Demonstrated</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Cumulative Program Ratings by Year (Initial Reviews Only)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective</td>
</tr>
<tr>
<td>Moderately Effective</td>
</tr>
<tr>
<td>Adequate</td>
</tr>
<tr>
<td>Ineffective</td>
</tr>
<tr>
<td>Results Not Demonstrated</td>
</tr>
</tbody>
</table>

When President George W. Bush created the PART program, he intended for each program to improve continually, increasingly making programs more accountable and better performing. To measure the continued improvement of programs, OMB places the burden on program managers to demonstrate results and drive improvement from the recommendations established under the original review.
Typically, programs were reassessed at the request of program managers, when they believed significant changes were likely to improve the rating of the program. For example, programs might be reassessed when new performance measures are implemented, follow-up actions have been completed, new performance data have been compiled, or a program evaluation has been completed.

Towards the end of the Bush Administration, the number of programs receiving their initial review significantly declined. As a result of the early emphasis for program reviews, few were left to receive their initial review. Additionally, the number of programs undergoing the re-review process also declined during this same period.

Table 3 and Chart 1 show the breakdown of program re-reviews in terms of change in magnitude for the 312 re-reviews conducted.

<table>
<thead>
<tr>
<th>Magnitude Change</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>86</td>
<td>46</td>
<td>83</td>
<td>54</td>
<td>23</td>
</tr>
<tr>
<td>n = 312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86</td>
<td>46</td>
<td>83</td>
<td>54</td>
<td>23</td>
</tr>
<tr>
<td>Mode = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86</td>
<td>46</td>
<td>83</td>
<td>54</td>
<td>23</td>
</tr>
<tr>
<td>Median = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86</td>
<td>46</td>
<td>83</td>
<td>54</td>
<td>23</td>
</tr>
<tr>
<td>Mean = 1.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86</td>
<td>46</td>
<td>83</td>
<td>54</td>
<td>23</td>
</tr>
</tbody>
</table>
Of the 312 reassessments, 6.4% (20) had a decrease in their PART score resulting from the re-reviews, 27.6% (86) had no change, while 66.0% (206) had an increase in their PART score. The median magnitude increase for a program requesting a reassessment was two, meaning that a program, for example, with an initial review rating of Adequate on average would increase to Effective. This weighting towards improvement is expected, as programs initiate the re-review process and programs are likely only going to request a reassessment if they believe their score improved.
Chapter 4: Results and Discussion

Model 1

Prior to running the initial regression for model 1, it became apparent that not all of the 1,015 programs contained all of the necessary information – primarily the dependent variable information, percentage change in budgeted amounts. Unfortunately, OMB does not track budget information by PART program overtime. Therefore, to complete the dataset, I needed to combine multiple spreadsheets of data. As part of this combination, 43 programs lacked the necessary budget information (budget amounts in the year prior and post review) and were subsequently dropped from the regression data. Additionally, OMB was unable to provide preliminary financial information for programs reviewed during the first year. Furthermore, the William D. Ford Direct Student Loan Program was dropped from the dataset because it had negative budget amounts for its relevant fiscal years, which would skew the results. Finally, three programs had classified budget during their relevant fiscal years and were therefore dropped. Because of the missing or incorrect information a total of 243 observations were marked as missing from the dataset, leaving 772 valid data points.
Initial Results

Table 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Error</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.23536</td>
<td>.10673</td>
<td>.0277**</td>
</tr>
<tr>
<td>Moderate</td>
<td>-.09495</td>
<td>.05384</td>
<td>.0782***</td>
</tr>
<tr>
<td>Adequate</td>
<td>-.10247</td>
<td>.05675</td>
<td>.0714***</td>
</tr>
<tr>
<td>Ineffective</td>
<td>-.24046</td>
<td>.13500</td>
<td>.0753***</td>
</tr>
<tr>
<td>RND</td>
<td>-.10653</td>
<td>.05702</td>
<td>.0621***</td>
</tr>
<tr>
<td>DirectFederal</td>
<td>.01929</td>
<td>.06301</td>
<td>.7595</td>
</tr>
<tr>
<td>Credit</td>
<td>-.08087</td>
<td>.10600</td>
<td>.4458</td>
</tr>
<tr>
<td>RandD</td>
<td>-.05327</td>
<td>.05893</td>
<td>.3663</td>
</tr>
<tr>
<td>BF_Grant</td>
<td>-.02606</td>
<td>.06517</td>
<td>.6893</td>
</tr>
<tr>
<td>CompGrant</td>
<td>-.07216</td>
<td>.05816</td>
<td>.2152</td>
</tr>
<tr>
<td>CapAssets</td>
<td>-.01617</td>
<td>.07222</td>
<td>.8229</td>
</tr>
<tr>
<td>Regulatory</td>
<td>-.00412</td>
<td>.07248</td>
<td>.9547</td>
</tr>
<tr>
<td>Political_Bias</td>
<td>-.03619</td>
<td>.04114</td>
<td>.3794</td>
</tr>
<tr>
<td>LnCurrentFY</td>
<td>-.01078</td>
<td>.00982</td>
<td>.2725</td>
</tr>
<tr>
<td>Agency_Size</td>
<td>.01295</td>
<td>.07067</td>
<td>.8547</td>
</tr>
</tbody>
</table>

\[ R^2 = .0223 \quad \text{Adjusted } R^2 = .0042 \]

* Statistically Significant at the 99% level
** Statistically Significant at the 95% level
*** Statistically Significant at the 90% level

Using Effective-rated programs as the omitted group, the OLS regression resulted in statistically significant results at the 90% level or greater for the key independent variables of interest related to the PART scores.

The results of the regression indicate that holding all other factors constant Effective-rated programs received an estimated average 23.536% percent increase in budget allocations. Moderately Effective-rated programs received an estimated average of 14.036 percent increase (.23536 - .09495); Adequate-rated programs
received an estimated average increase of 13.286; while, Ineffective-rated programs on average received an estimated allocation decrease of 0.514 percent. Programs rated as Results Not Demonstrated received an average estimated increase of 12.883 percent. Holding all other variables in the model constant, these results indicate that programs receiving lower rating levels on average received lower budget increases (illustrated in Chart 2).

Plotting these data points, my hypothesis that higher rated programs received larger increases is initially confirmed. The notable exception to this trend is programs rated as Results Not Demonstrated; these programs received average increases nearly on par with Adequate programs. Because RND ratings are given to programs where
managers have either not developed goals or not collected sufficient data, OMB occasionally refers to these programs as having ratings outside the core Effective to Ineffective scale. As such, it is not entirely surprising to see that RND programs are not penalized as Ineffective programs.

Additionally, these results are consistent with the findings of Gilmour and Lewis, where they found when accounting for similar control variables, PART scores showed a statistically significant and directional relationship within President Bush’s FY2004. In their study, Gilmour and Lewis also found that the apparent effect of PART scores on budgeted levels increased when just looking at programs housed within traditionally Democratic agencies (Department of Commerce, Education, Energy, Housing and Urban Development, Labor, Health and Human Services and the Environmental Protection Agency). Applying a similar criteria across the entire dataset of PART years, the results were less conclusive (See Table 5). For programs housed in Democratic agencies, absent the intercept, the results were entirely non-significant. Conversely, for programs in non-traditional Democratic agencies, several program rating had a significant relationship with the expected signs. These results indicate that PART scores seemed to have a more relevant impact on determining the budget allocations on programs in non-Democratic agencies during the Bush Presidency.
Table 5

Abbreviated Results: PART Scores Separated by Political Bias

<table>
<thead>
<tr>
<th>Variable</th>
<th>Non-Democratic Agencies (political_bias = 0)</th>
<th>Democratic Agencies (political_bias = 1)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.23212***</td>
<td>.22217**</td>
</tr>
<tr>
<td>Moderate</td>
<td>-.14678**</td>
<td>.04356</td>
</tr>
<tr>
<td>Adequate</td>
<td>-.12441</td>
<td>-.03649</td>
</tr>
<tr>
<td>Ineffective</td>
<td>-.42006***</td>
<td>-.10079</td>
</tr>
<tr>
<td>RND</td>
<td>-.08253</td>
<td>-.10685</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.0226</td>
<td>.0752</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>-.0038</td>
<td>Adjusted $R^2$ = .0332</td>
</tr>
<tr>
<td>Obs. Used</td>
<td>495</td>
<td>277</td>
</tr>
</tbody>
</table>

* Statistically Significant at the 99% level  
** Statistically Significant at the 95% level  
*** Statistically Significant at the 90% level  
**** All programs in traditionally Democratic Agencies were in large agencies; therefore, agency size was dropped from this regression.

While these findings differed from Gilmour and Lewis’ research, numerous empirical explanations exist, including that the likelihood that the emphasis and application of PART scores likely differed over time, creating varying results within each year. Additionally, the coefficients on Adequate and Results Not Demonstrated variables were not significant at traditional measures, creating uncertainty in interpreting these results.

**Accounting for Time**

One key concern with the regression results presented above is that empirical analysis of PART scores, illustrated in Chart 3, shows a trend towards improving overall scores. This trend is most noticeable by considering the percentage of programs rated as Results Not Demonstrated in 2002 (39%) versus 2008 (16%) and
programs rated as performing (Effective + Moderately Effective + Adequate) in 2002 (57%) versus 2008 (80%).

Following PART’s introduction in 2002, it is likely that many program managers were unfamiliar with the rating tool’s measurement techniques and goals. As a result, a higher percentage of programs did not have the necessary metrics to measure their goals. Overtime as managers became more familiar with PART, metrics were created and programs were able to receive ratings that better captured a program’s success. However, as PART continued, many managers likely realized, whether real or perceived, that a better PART score could improve the appearance of their program. As a result, shrewd managers could create less challenging goals,
allowing their programs to succeed more easily and achieve a higher rating. This method of ‘gaming the system,’ creates a problem similar to grade inflation in schools, where later grades likely have less meaning given the upward skew. Grade inflation among PART scores could likely result in high scores in later years having less relevance in budget decisions as more and more programs are rated as performing. As a result, PART scores in later years would be expected to have less significant results, particularly among performing programs. However, running the regression yields different results. Table 6 contains the regression results by adding review years into the model – calendar year 2008 and programs rated as Effective serve as the omitted categories.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Error</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.34145</td>
<td>.14879</td>
<td>.0220 **</td>
</tr>
<tr>
<td>Moderate</td>
<td>-.09038</td>
<td>.05404</td>
<td>.0949***</td>
</tr>
<tr>
<td>Adequate</td>
<td>-.10141</td>
<td>.05705</td>
<td>.0759***</td>
</tr>
<tr>
<td>Ineffective</td>
<td>-.23923</td>
<td>.13575</td>
<td>.0784***</td>
</tr>
<tr>
<td>Results Not Demonstrated</td>
<td>-.10773</td>
<td>.05800</td>
<td>.0636***</td>
</tr>
<tr>
<td>CY2003</td>
<td>-.14621</td>
<td>.12009</td>
<td>.2238</td>
</tr>
<tr>
<td>CY2004</td>
<td>-.08798</td>
<td>.11836</td>
<td>.4575</td>
</tr>
<tr>
<td>CY2005</td>
<td>-.17175</td>
<td>.11872</td>
<td>.1484</td>
</tr>
<tr>
<td>CY2006</td>
<td>-.14434</td>
<td>.11849</td>
<td>.2235</td>
</tr>
<tr>
<td>CY2007</td>
<td>-.09354</td>
<td>.13318</td>
<td>.4827</td>
</tr>
<tr>
<td>DirectFederal</td>
<td>.02524</td>
<td>.06354</td>
<td>.6913</td>
</tr>
<tr>
<td>Credit</td>
<td>-.08430</td>
<td>.10647</td>
<td>.4287</td>
</tr>
<tr>
<td>RandD</td>
<td>-.05103</td>
<td>.05904</td>
<td>.3876</td>
</tr>
<tr>
<td>BF_Grant</td>
<td>-.01675</td>
<td>.06543</td>
<td>.7980</td>
</tr>
<tr>
<td>CompGrant</td>
<td>-.07107</td>
<td>.05830</td>
<td>.2232</td>
</tr>
<tr>
<td>CapAssets</td>
<td>-.01033</td>
<td>.07245</td>
<td>.8866</td>
</tr>
</tbody>
</table>
Adding indicator variables representing the calendar year during which a program received its initial review, the overall outcome of the models bears similarities to the initial model. The same relationship between PART scores and budget allocations still exists – higher rated programs, holding all else constant, on average receive higher budget increases, and RND-rated programs still average estimated increases slightly below Adequate-rated programs, but above Ineffective programs. These results indicate that adding variables to control for year of review does not yield drastically different results. Additionally, none of the coefficients on the year variables were statistically significant at traditional levels, indicating that the year a program was reviewed is potentially irrelevant.

Re-evaluating the grade inflation concern, as more programs are rated higher the benefit of a higher score should mean less; however, Ineffective and Results Not Demonstrated ratings will become more rare and therefore should have a greater negative impact on budgeted levels of later reviewed programs. To better evaluate this notion, I created interaction variables between PART scores and the year of review. Comparing the results of the interactions provides a more appropriate year-to-year
comparison. Table 7 contains the result of the interactive model regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Error</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.89181</td>
<td>.21886</td>
<td>.0001*</td>
</tr>
<tr>
<td>Moderate</td>
<td>-.87413</td>
<td>.34476</td>
<td>.0114**</td>
</tr>
<tr>
<td>Adequate</td>
<td>-.96929</td>
<td>.27897</td>
<td>.0050*</td>
</tr>
<tr>
<td>Ineffective</td>
<td>-.76320</td>
<td>.52510</td>
<td>.1465</td>
</tr>
<tr>
<td>Results Not Demonstrated</td>
<td>-.83529</td>
<td>.34437</td>
<td>.0155**</td>
</tr>
<tr>
<td>CY2003</td>
<td>-.62352</td>
<td>.22756</td>
<td>.0063*</td>
</tr>
<tr>
<td>CY2004</td>
<td>-.77812</td>
<td>.2211</td>
<td>.0005*</td>
</tr>
<tr>
<td>CY2005</td>
<td>-.75887</td>
<td>.21508</td>
<td>.0004*</td>
</tr>
<tr>
<td>CY2006</td>
<td>-.54979</td>
<td>.21524</td>
<td>.0108**</td>
</tr>
<tr>
<td>CY2007</td>
<td>-.93033</td>
<td>.23575</td>
<td>.0001*</td>
</tr>
<tr>
<td>DirectFederal</td>
<td>.01991</td>
<td>.06325</td>
<td>.7530</td>
</tr>
<tr>
<td>Credit</td>
<td>-.09458</td>
<td>.10630</td>
<td>.3739</td>
</tr>
<tr>
<td>RandD</td>
<td>-.05367</td>
<td>.05937</td>
<td>.3663</td>
</tr>
<tr>
<td>BF_Grant</td>
<td>-.02139</td>
<td>.06538</td>
<td>.7437</td>
</tr>
<tr>
<td>CompGrant</td>
<td>-.07339</td>
<td>.05825</td>
<td>.2082</td>
</tr>
<tr>
<td>CapAssets</td>
<td>-.01003</td>
<td>.07269</td>
<td>.8903</td>
</tr>
<tr>
<td>Regulatory</td>
<td></td>
<td>.07286</td>
<td>.8632</td>
</tr>
<tr>
<td>Political_Bias</td>
<td>-.03980</td>
<td>.04151</td>
<td>.3380</td>
</tr>
<tr>
<td>LnCurrentFY</td>
<td>-.00656</td>
<td>.00995</td>
<td>.5104</td>
</tr>
<tr>
<td>Agency_Size</td>
<td>.00041</td>
<td>.07259</td>
<td>.9954</td>
</tr>
<tr>
<td>MOD2003</td>
<td>.72490</td>
<td>.37142</td>
<td>.0514***</td>
</tr>
<tr>
<td>MOD2004</td>
<td>.91920</td>
<td>.36352</td>
<td>.0117**</td>
</tr>
<tr>
<td>MOD2005</td>
<td>.88998</td>
<td>.35967</td>
<td>.0136**</td>
</tr>
<tr>
<td>MOD2006</td>
<td>.56573</td>
<td>.36077</td>
<td>.1173</td>
</tr>
<tr>
<td>MOD2007</td>
<td>1.31313</td>
<td>.38922</td>
<td>.0008*</td>
</tr>
<tr>
<td>ADQ2003</td>
<td>.076495</td>
<td>.13417</td>
<td>.0151**</td>
</tr>
<tr>
<td>ADQ2004</td>
<td>1.11023</td>
<td>.30299</td>
<td>.0003*</td>
</tr>
<tr>
<td>ADQ2005</td>
<td>.96631</td>
<td>.29864</td>
<td>.0013*</td>
</tr>
<tr>
<td>ADQ2006</td>
<td>.66560</td>
<td>.30019</td>
<td>.0269**</td>
</tr>
<tr>
<td>ADQ2007</td>
<td>1.08647</td>
<td>.34802</td>
<td>.0019*</td>
</tr>
<tr>
<td>INE2003</td>
<td>.48029</td>
<td>.57285</td>
<td>.4021</td>
</tr>
<tr>
<td>INE2004</td>
<td>.73831</td>
<td>.60410</td>
<td>.2220</td>
</tr>
<tr>
<td>INE2005</td>
<td>.38130</td>
<td>.60025</td>
<td>.5255</td>
</tr>
<tr>
<td>Year</td>
<td>Coefficient 1</td>
<td>Coefficient 2</td>
<td>Coefficient 3</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>INE2006</td>
<td>0.54209</td>
<td>0.63468</td>
<td>0.3933</td>
</tr>
<tr>
<td>RND2003</td>
<td>0.66303</td>
<td>0.36681</td>
<td>0.0711***</td>
</tr>
<tr>
<td>RND2004</td>
<td>0.8808</td>
<td>0.36209</td>
<td>0.0152 **</td>
</tr>
<tr>
<td>RND2005</td>
<td>0.64787</td>
<td>0.36512</td>
<td>0.0764***</td>
</tr>
<tr>
<td>RND2006</td>
<td>0.67192</td>
<td>0.36236</td>
<td>0.0641***</td>
</tr>
<tr>
<td>RND2007</td>
<td>1.17994</td>
<td>0.39645</td>
<td>0.0030*</td>
</tr>
</tbody>
</table>

$R^2 = 0.0751$  
Adjusted $R^2 = 0.0272$

* Statistically Significant at the 99% level
** Statistically Significant at the 95% level
*** Statistically Significant at the 90% level

INE2007 was dropped from the model, because no program received an Ineffective rating in 2007

Using Effective-rated programs reviewed in calendar year 2008 as the omitted group, the OLS regression resulted in statistically significant results at the 90% level or greater for nearly all of the key independent variables of interest related to the PART scores. Table 8 translates the regression interactive term results into coefficient to more easily illustrate the results. These numbers are calculated by adding the base case coefficient with all portions of the interactive terms. For example, the coefficient for programs rated as Moderate in 2004 is calculated by adding the Base Case coefficient (0.89181) with the Moderate coefficient (-0.87143), the 2004 coefficient term (-0.77812), and the interactive term (0.91920). The final number indicates that holding all else constant, programs rated as Moderate in 2004 received an average budget increase of 15.876 percent.
An important limitation of the interactive model is that the coefficients on all Ineffective-rated programs contained at least one input that was not statistically significant at conventional levels; as a result the corresponding coefficients are difficult to interpret in the existing model.

Another interpretative challenge is that in calendar years 2007 and 2008, the number of program reviews significantly dropped. This sharp decrease places greater emphasis on the fewer remaining programs in the analysis, making them more susceptible to outliers. For example, in calendar year 2008 OMB reviewed 19 programs, rating six as Effective; of these programs, two received extremely large (over 200%) allocation increases. These programs, the Federal Deposit Insurance Corporation’s (FDIC) Regulation and Examination program and the FDIC’s Deposit Insurance Fund, received large capital inflows as their operational scope and responsibilities widely expanded in response to the financial crisis. Due to this dramatic shift, it is difficult to judge the overall effect of these two program’s PART
rating. However, as a measure of comparison, the remaining four programs rated as Effective in 2008 received an average allocation decrease of 5.850 percent. As a result, outliers may more heavily bias the results in calendar years 2008 and 2007.

Plotting the results from the interactive model by rating, the hypothesis of grade inflation-style effects appears incorrect (See Chart 4). The grade inflation hypothesis theorized that high-performing grades would depreciate in value overtime, resulting in decreased relative worth; while low-performing and RND results would become more detrimental in value because of their relative rarity. However, across all years rating levels there is no relevant pattern or movement.

Chart 5 illustrates the same interactive model results plotted by year; showing how ratings may effect budget allocation in a given year. In this chart, the initial year
(2003) appeared as hypothesized – higher ratings, holding all else constant, typically result in higher budget levels. For calendar year 2004, the results appear generally inversed; where Moderate-rated programs, holding all else constant, received higher allocation increases than any other rating. In calendar year 2005, all reviewed performing programs received nearly identical budget increase of 12-14%, while Ineffective and RND-rated programs received substantially less. For 2005, these result indicate that PART’s usage could have been applied as a basic filter to weigh programs as either performing or not performing. Programs reviewed in calendar year 2006 appear to have inverse relationship with the exception of Effective-rated programs. The results for programs reviewed in 2007, fail to illustrate a well-defined trend, with Moderate and RND-rated programs receiving significantly higher allocation increases. These results, as well as those for calendar year 2008 are potentially more skewed by outliers, based on the comparatively fewer observations.
The variances illustrated in Charts 4 and 5, could have resulted if OMB varied its usage of PART each year. I believe this theory to be unlikely, as changes in application of PART appear drastic and often inconsistent with its purpose. A more likely scenario is that PART is a tool with little or no consistent usage in the overall budget process. Although I attempted to account for several other variables that affect the President’s budget, an omitted bias likely still pervades the data. The measurement of political bias, while expedient, is fairly basic and the models fail to fully capture political trends (such as monetary support of the financial industry caused by recession), among a host of other potential concerns.

The results of the simple model regression analysis illustrate a significant relationship between PART scores and budget allocation increases in the President’s
budget. However, after creating a more complex model that allows for interaction between the year of review and the rating, this relationship is significant, but inconsistent over time.

Furthermore, PART, from its initiation was created as a tool, not a decision mechanism. While these results do not necessarily illustrate a consistent pattern, this lack of a trend is not entirely unexpected as the Bush Administration claimed PART’s true effectiveness was just advisory.

The results from these models certainly question the useful of PART as a continued budgeting tool. If PART continues to assist OMB and program managers locate areas of improvement, it could still prove an effective management tool. However, without greater application of PART results in the White House budget process, PART’s effectiveness as a budget tool is certainly called into question.

**Model 2**

In the varieties Model 1, I only considered the effects initial reviews had on budget allocation increases. However, throughout President Bush’s term, 262 program were re-reviewed (49 of these programs were re-reviewed two or more times, creating a dataset of 312 observations). The relationship between re-reviewed ratings and budget allocation increases is important because, it could illustrate if programs that undergo the re-review process actually gain financially. The regression results for Model 2 are displayed in Table 9.
### Table 9

<table>
<thead>
<tr>
<th>Variable</th>
<th>Base Model Coefficient Value</th>
<th>Base + Prev. Results Coefficient Value</th>
<th>Interactive Model Coefficient Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.17492</td>
<td>1.68402</td>
<td>1.51612</td>
</tr>
<tr>
<td>Magnitude</td>
<td>-.10343</td>
<td>-.11271</td>
<td>-.08295</td>
</tr>
<tr>
<td>DirectFederal</td>
<td>-.03436</td>
<td>-.02618</td>
<td>-.06546</td>
</tr>
<tr>
<td>Credit</td>
<td>-.47821</td>
<td>-.46547</td>
<td>-.58778</td>
</tr>
<tr>
<td>RandD</td>
<td>-.48429</td>
<td>-.46382</td>
<td>-.48522</td>
</tr>
<tr>
<td>BF_Grant</td>
<td>-.16506</td>
<td>-.15986</td>
<td>-.19831</td>
</tr>
<tr>
<td>CompGrant</td>
<td>-.59695</td>
<td>-.59257</td>
<td>-.54436</td>
</tr>
<tr>
<td>CapAssets</td>
<td>1.05098</td>
<td>1.06488</td>
<td>1.01153</td>
</tr>
<tr>
<td>Regulatory</td>
<td>-.75544</td>
<td>-.73562</td>
<td>-.70525</td>
</tr>
<tr>
<td>Political_Bias</td>
<td>-.32296</td>
<td>-.31972</td>
<td>-.37476</td>
</tr>
<tr>
<td>LnCurrentFY</td>
<td>-.33740*</td>
<td>-.33651*</td>
<td>-.33846*</td>
</tr>
<tr>
<td>Agency_Size</td>
<td>.26030</td>
<td>.38772</td>
<td>.36142</td>
</tr>
<tr>
<td>Span</td>
<td>.06004</td>
<td>.04048</td>
<td>.16188</td>
</tr>
<tr>
<td>PrevModerate</td>
<td>.57899</td>
<td>.31320</td>
<td></td>
</tr>
<tr>
<td>PrevAdequate</td>
<td>.26564</td>
<td>.67708</td>
<td></td>
</tr>
<tr>
<td>PrevIneffective</td>
<td>.44059</td>
<td>.46582</td>
<td></td>
</tr>
<tr>
<td>PrevRND</td>
<td>.51251</td>
<td>1.00079</td>
<td></td>
</tr>
<tr>
<td>SMod</td>
<td>.08510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAdeq</td>
<td>.019423</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIne</td>
<td>-.02761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRND</td>
<td>-.24031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.0512</td>
<td>.0543</td>
<td>.0579</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.0124</td>
<td>.0021</td>
<td>-.0080</td>
</tr>
</tbody>
</table>

* Statistically Significant at the 99% level
** Statistically Significant at the 95% level
*** Statistically Significant at the 90% level

In these tested situations, the predominance of variables were not statistically significant, therefore not proving a definitive relationship between PART score improvements from re-reviews and allocation increases. The initial version of Model 2 is testing to see if there is a basic relationship between the change of a program’s
PART rating and the percentage change in budget the following year. The results from this initial model indicate that the best predictor of future budget increase (and the only significant result) is a program’s previous budget level.

The second version of this model, which incorporates a program’s previous review score, yields similar results. Additionally, the third version of this model, which includes the interaction between the time variable (span) and the previous results ratings, yields nearly identical results.

A potential explanation for these inconclusive results is multi-faceted, including several key notions:

1. The lack of a consistent relationship between initial PART ratings and allocation increases, as illustrated in the original interactive model, remains for re-reviews. The results from the interactive version of Model 1 (Table 8) showed an inconsistent use of the original rating. I would expect the results from the re-reviews to suffer the same shifting relationship in use.

2. The self-selecting nature of the re-review process creates a bias where the vast majority of programs only undergo a re-review if they believe they can enhance their rating level. If a link between PART ratings and budget increases existed (or program managers thought there was a link), managers would only seek to dedicate resources to the re-review process if they knew an increase in ratings was
achievable. As a result, a majority of re-reviews resulted in increases (about 66%), limiting the variation in the model and creating interpretation difficulties.

3. Increases in funding, irrelevant of PART, may actually increase program re-review ratings. The notion of this thesis is that PART results potentially drive budget increases; however, it is possible that the opposite relationship exists, where increases in budgets between re-reviews could allow for improved PART scores. In this circumstance, the regressions would likely show inconclusive results.

Combined, these factors illustrate little funding level value for program managers to undergo the re-review process. However, because several programs managers were quoted in government-oriented media publications discussing how PART helped their program create internal metric and improvement plans, the PART re-review process still held some potential benefit for programs.
Chapter 5: Conclusion and Policy Implications

The initial results of this thesis provide evidence that PART reviews are a contributing factor in percentage changes of allocations in the President’s Budget. However, the interactive model, which accounts for change overtime, appears contradictory with inconsistent results in multiple years. This contradiction creates difficulty in determining the overall policy implications; however, several options may increase the overall usefulness of PART.

1. Budgeting is an inherently political process that reflects a wide-range of competing needs and interests. However, PART strives to be a non-partisan tool within a highly partisan environment; as a result, this thesis shows it usage is mixed depending on the model used. Tat-Kei Ho points out that a performance management system must measure the items decision makers are interested in to be useful. In the case of the budget, it may be necessary for a performance manage tool to include partisan measurements and goals. While White House inclusion of partisan items will likely exclude Congressional Republican use of the tool, it would create a more effective tool of OMB. As shown by the interactive version of Model 1 and the re-review results of Model 2, President Bush’s use of PART as a tool was likely sporadic and may have created confusion regarding its overall benefit to the Administration.

Example political inclusions for President Obama could include

51
measurements as to whether a program is in line with current Administration priorities, such as promoting job growth and stability, or whether the program has taken steps to become more environmentally efficient.

2. The Executive Branch with little Congressional input created PART; conversations between agencies and OMB in the budget setting process are not public information, excluding Congress from this important knowledge; and Congress creates the final budget allocations. Combined, these three situations create a dilemma where the information that is available is not necessarily what final decision makers would ideally want or use to make performance management-driven budget decisions. While it is unlikely that President Obama (or any of his predecessors) will share information gather during the preliminary discussions between OMB and the agencies, PART can be improved by re-tooling the process to include evaluations and measurements important to Congress. Updating (or replacing) PART with a performance management tool that meets the demands of both the White House and Congress, creates a universal and non-partisan tool that all parties can use to begin debate.

While these two notions appear contradictory, the bottom line is that performance management systems need to generate relevant results to be useful. Previous government systems faded when the politics of the President changed or we
changed Presidents; and it is likely that we will continue to see seemingly continuous political changes. PART’s usage was particularly strong for programs reviewed in 2003, but, as the interactive version of Model 1 shows, its effectiveness rapidly diminished. As a result, policymakers would be better off creating an effective performance management system that is adaptable to meet the dynamics of its users – allowing it to stay relevant and informative.

Limitations and Direction for Future Research

Although the methods utilized in this study produced statistically significant results (in Model 1), the analysis does have limitations. In particular, the public availability of information budget information was constraining. One key measure of interest in the budgeting process was the ratio between what individual programs and agencies requested for budget increases versus what was actually received in the President’s Budget. Unfortunately, budget conversations between agencies and OMB are not made public. In many ways, this ratio would be a stronger test for PART’s effectiveness, as it could help explain in a deeper fashion the decision making process. For example, an Effected-rated program could request and receive in entirety a very small increase to cover cost-of-living increases; whereas, an Ineffective program could request an extremely large percentage increase, but only receive a modest percentage. Under the current regression model, these results would create confusion as to PART’s effectiveness (as done in the interactive version of Model 1). A regression model
including the knowledge of original requests, should be able to help explain additional variation within the model by analyzing the true decision making step.

An additional notion of further consideration is to research the relationship between PART scores to Congressional Budget allocations. Because Congress has the final budget allocation authority (absent the President’s veto power), their use of a performance management system would clearly justify the costs the government incurs through PART’s implementation. Unfortunately, OMB’s decisions regarding the separation of government services into the 1,015 programs is proprietarily used by the White House. As a result, the process to align these programs with the final budget allocations distributed by Congress would require an overly cumbersome amount of research – making the analysis extremely difficult, if not impossible.
Endnotes:
1. Adam Hughes, Director of Federal Fiscal Policy at the nonprofit government-watchdog group OMB Watch, as quoted by: Newell, Elizabeth. See Newell, E 2009.
2. See Georgetown Public Policy Institute, Accenture Consulting and OMB Watch 2009
3. See OMB 2001
4. In its 2006 review of the PART, the US GAO noted that OMB’s PART recommendations increased program manager’s attention, focus and resource investment in evaluation.
6. See Gruber February 3, 2003
7. See House Committee on Government Reform, Subcommittee on Government Efficiency and Financial Management 2004
8. This is just an example of many references to articles that state the benefits to those programs that utilize the PART, but also indicate that far from every program utilizes the program - See Mosquera 2008
9. Michael Stephens, a senior staff assistant on the House Appropriations Committee, as quoted by Gruber, Amelia June 13, 2003. While this is one instance, many other quotes suggest similar opinions.
10. See Gilmour and Lewis: *Does Performance Budgeting Work? An Examination of the Office of Management and Budget’s PART Scores*,

11. See Gilmour and Lewis: *Assessing Performance Budgeting at OMB: The Influence of Politics, Performance, and Program Size*

Bibliography and Referenced Documents:


• Gruber, A. (February 3, 2003). Poor Performance Leads to Budget Cuts at Some


• President’s Budget for the Fiscal Year 2004: Rating the Performance of Federal Programs, p. 53.


• US Congress, House Committee on Government Reform, Subcommittee on Government Efficiency and Financial Management (February 11, 2004). The


