External and Internal Agency Characteristics as Determinants of Program Assessment
Rating Tool (PART) Scores

by Jennifer Deets Sayasithsena

B.A. 1997, Georgetown University
M.A. 1999, University of Southern California

A Dissertation Submitted to

The Faculty of

Columbian College of Arts and Sciences
of the George Washington University in partial satisfaction
of the requirements for the degree of Doctor of Philosophy

August 31, 2010

Dissertation directed by

Steven J. Balla
Associate Professor of Political Science, of Public Policy and Public Administration,
and of International Affairs

The Columbian School of Arts and Sciences of The George Washington University

certifies that Jennifer Deets Sayasithsena has passed the Final Examination for the degree

of Doctor of Philosophy as of July 23, 2010. This is the final and approved form of the

dissertation.

External and Internal Agency Characteristics as Determinants of Program Assessment
Rating Tool (PART) Scores

Jennifer Deets Sayasithsena

Dissertation Research Committee:

Steven J. Balla, Associate Professor of Political Science, of Public Policy and Public
Administration, and of International Affairs, Dissertation Director

Paul J. Wahlbeck, Professor of Political Science, Committee Member

Robert P. Stoker, Associate Professor of Political Science and of Public Policy and Public
Administration, Committee Member

Dedication

I wish to dedicate this dissertation to my parents, Horace and Connie Deets, for their incredible support of my academic pursuits and in all aspects of my life. I also dedicate this dissertation to my wonderful husband, Benny, and our daughters Madeline and Natalia, for their support and patience during my many years in graduate school.

Acknowledgements

I acknowledge the tremendous time dedicated to this dissertation by Steven Balla, Paul

Wahlbeck and Bob Stoker as they read, made suggestions and commented on many

versions of this document.

Dissertation Abstract

EXTERNAL AND INTERNAL AGENCY CHARACTERISTICS AS DETERMINANTS OF PROGRAM ASSESSMENT RATING TOOL (PART) SCORES

In this dissertation I explore why some federal agency programs received better Program Assessment Rating Tool (PART) scores than others during the Bush administration. PART scores served as the primary measure of agency program performance during the Bush administration. Principal-agency theory suggests that the president and congress use mechanisms of external control such as budgeting and appointments to constrain agency behavior and influence performance. On the other hand, alternate theories suggest internal agency characteristics, including leadership and ideology, are responsible for performance.

I first review the long history of performance management at the federal level. I then use Program Assessment Rating Tool (PART) scores implemented during the Bush administration to empirically examine federal agency program performance. More than 1,000 PART scores from 2002-2008 are examined in separate models to test the effects of external, internal and combined characteristics of agency program performance. I find that both internal and external characteristics matter and that agency politicization and leadership are especially important predictors of performance on PART.

Following my empirical analysis, I explore agency culture as an additional explanation of agency performance beyond PART. Culture is thought to be very important, yet is nearly impossible to quantify. I use several agencies as exploratory case studies to further examine culture's impact on performance. In my concluding chapter I highlight the Obama administration's current performance management efforts.

Table of Contents

List of Tables

List of Tables, cont.

CHAPTER ONE

Introduction and Literature Review

Introduction to the Research Question:

Government agencies provide a wide range of services to many different constituencies with varying degrees of success. Often, government agencies serve people classified as "at risk" including the homeless, elderly, and runaway youth. Due to the national attention concerning runaway and homeless youth in the 1970s, Congress passed the Juvenile Justice and Delinquency Prevention Act (JJDPA) of 1974. The JJDPA provided the infrastructure for a system of care for young runaways outside the traditional law enforcement and juvenile justice agencies. Through the JJDPA's Runaway and Homeless Youth Act (RHYA), Congress authorized the Basic Center Program to be housed in the Department of Health and Human Services (DHHS). The program, administered by the DHHS' Family and Youth Services Bureau, funds more than 330 local community programs serving runaway and homeless youth. The program provides food, shelter, clothing, counseling and health care referrals and operates on an annual budget of $96 million (FY 2008) (ExpectMore.gov). The Basic Center program is classified as a competitive grant program.

The HOPE for Elderly Independent (HOPE IV) program administered by the Department of Housing and Urban Development also serves an at-risk constituency. The HOPE IV program provides funding through grants to state, county and municipal government agencies to operate or develop housing assistance programs for the elderly and disabled. The program provides housing, food, health care and counseling to low-

income elderly and disabled to prevent them from premature placement in nursing homes. The HOPE IV program operates on an annual budget of $100 million (FY 2008) and is also classified as a competitive grant program (ExpectMore.gov).

The two programs, the Basic Center Program and HOPE IV, share several key characteristics. The annual operating budgets are similar; both are administered as competitive grant programs housed in cabinet-level agencies, and both serve at-risk populations providing housing, food, health care and counseling services. There is, however, a notable difference. The Basic Center Program received stellar performance scores on the Office of Management and Budget's (OMB) Program Assessment Rating Tool (PART) in all four rating categories: program purpose and design, planning, management and results. The program's overall rating is "excellent." According to OMB, "excellent" programs, "set ambitious goals, achieve results, are well-managed and improve efficiency." The HOPE IV program, on the other hand, received an overall score of "ineffective," which means the program has "been unable to achieve results due to a lack of clarity regarding the program's purpose or goals, poor management, or some other significant weakness." The two programs face similar challenges. Both provide services to at-risk constituencies, operate on budgets that are roughly the same size, and are of the same funding type. But they performed quite differently on PART. Other similar programs likely received different PART scores too, but why? What explains such variation in PART scores? Are the inner workings of an agency responsible? Might external factors such as structure and divided government matter? There are many possible explanations and I will revisit these two programs later in this dissertation (see Chapter 3).

This dissertation explores why some agency programs performed better than others on PART. There are few large-N quantitative studies of agency performance (notable exceptions include Krause and Douglas 2005; Meier and O'Toole 2006; Lewis 2008). This dissertation will use Program Assessment Rating Tool (PART) scores for more than 1,000 agency programs. The PART scoring system, implemented in 2003 by the Bush administration, provides not only an overall performance score which serves as the dependent variable, but also measures factors including agency budget size and program type. Case studies will be conducted in addition to the quantitative analysis. Cases studies are a valuable tool to further explore explanatory variables like agency culture and mission, which are often difficult to quantify. The case studies will be used to learn more about how culture operates in different agencies. Particular case study selection criteria are examined further in chapter 4.

Chapter Preview:

In this chapter I first review alternate theories of agency performance, which provide the rationale for the structure of my empirical analysis in Chapter 3. Second, I examine the evolution of performance management and measurement efforts and how these efforts have shaped decision-making at the federal level. Third, I discuss the many challenges associated with performance measurement. Fourth, I introduce the dependent variable for my dissertation, Program Assessment Rating Tool (PART) scores, which were developed and implemented during the Bush administration. Following a detailed description of the scores and how they work, I discuss both the shortcomings and impact of PART.

LITERATURE REVIEW

Explanations of Agency Performance:

How well bureaucratic agencies perform is important for a number of reasons. The primary theoretical reason is that agencies are made up of unelected officials who make important, binding policy decisions on a daily basis and there is a need to determine how accountable these agencies are to the public and other branches of government (Furlong 1998). A majority of federal government programs are now administered by third parties who are under contract, which makes accountability even more important (Frederickson and Frederickson 2007). At a more practical level, the public, the president and congress want to make sure that agencies are spending their budgets wisely and producing quality outcomes. The federal government spends more than $2 trillion annually on roughly 1,000 federal programs and yet, in most cases, there is no way to measure how well that money is spent (OMB 2004). The president and congress are also concerned with how agencies perform because program failures and successes may affect their own electoral fortunes.

For the public, trust in government has declined for decades. Strong federal agency performance could play an important role in reversing this trend. According to the Partnership for Public Service and Gallup Consulting's 2008 poll, "In the Public We Trust,"

> **Americans consistently rate government performance higher when they believe government gets the process right—by spending money wisely, making decisions based on merit and not politics, hiring talented civil servants, and demonstrating strong leadership.**

The different explanations for agency performance can be organized into two broad categories. The first is that agencies have autonomy and are able to withstand the

president and congress' control mechanisms because of internal characteristics like strong leadership, constituency support, and "agency culture" (Furlong 1998). It is internal characteristics as such that influence agency behavior and performance. The second explanation is that institutions outside the bureaucracy, specifically congress and the president, influence agency behavior in what is called a principal agent model. In this model, external political control mechanisms wielded by congress and the president are important indicators of outcomes and performance.

The president and congress provide incentives to agencies to get them to carry out their policy preferences. The president and congress dictate their preferences from the "top" to the agencies at the "bottom." The problem with this model, according to Meier and O'Toole, is that no "top" exists in our separation of powers system. While they agree that external mechanisms of control wielded by Congress and the president play an important role in how agencies perform, they are only *part* of the explanation. This approach ignores the fact that agencies have preferences too, such as a commitment to democratic norms and a desire to do a good job (Brehm and Gates 1997). These internal agency preferences might help agencies withstand control mechanisms employed by the president and congress. As Matthew Dull (2006) states, "while inter-organizational control remains the focus of most political science research on administrative politics, intra-organizational arrangements, including organizational and professional norms, play a decisive role in democratic control of public agencies." These intra-organizational or internal characteristics might be as important to dictating agency behavior and performance as external characteristics such as agency design and budget control.

Why is this important? If internal agency characteristics are as reliable predictors of agency performance as external characteristics, it could undermine the focus by many political scientists on principal agent theory so often used to explain the relationship between executive agencies, congress and the president. This then limits the extent to which the president and congress are the sole sources of influence on agency behavior and performance. Other factors, like culture and other internal characteristics, might play an important role.

Each of these explanations will be explored in detail in the next chapter. There are numerous studies that separately test and confirm or dismiss the link of external characteristics with agency performance (Moe 1982, 1985; Wood 1988, 1990; Wood and Waterman 1991, 1993; Carpenter, 1996; Krause 1996; Scholz and Wood 1999; Balla and Wright 2001). However, I argue that internal agency characteristics are often excluded from empirical studies of performance and are important predictors of agency performance. I draw upon findings from studies of organizational culture and institutional development (Wilson 1989, Hall 1996, Carpenter 2001). By excluding such important variables, previous analyses have potentially underspecified models leading to incorrect conclusions about what exactly explains agency performance.

The Evolution of Performance Measurement

Performance in the context of the federal bureaucracy is a concept that poses great difficulty for political scientists, policy analysts and public administration scholars alike. It is much easier to measure the performance of agencies with clearly observable outputs and outcomes. While businesses measure performance through well-established market

indicators like stock values, bureaucratic performance is more difficult to assess because of the variance in size, services they provide and the "clients" they serve.

For bureaucracies generally, performance can be measured in a number of ways. Outputs are the actions over which agencies exert direct control (e.g., the number of inspections carried out), while outcomes are the results agencies seek to bring about (e.g., cleaner air). Also important to consider is the performance context. Do agencies perform well day-to-day or in crisis situations? Are there differences between policymaking and implementation? Is performance synonymous with efficiency, or are there other dimensions that are important (Balla and Gormley 2003).

It is important to note that political actors have multiple and even competing goals (Baum 1997). For the president and congress, policy outcomes are especially important. Outcomes and performance are not necessarily synonymous, which is why it is necessary to draw some distinctions between the two concepts and properly define performance.

In the public sector, "performance" has increasingly become synonymous with results. Outcomes are valued above inputs, processes, or outputs (Radin 2000). Achieving stated outcomes is the definition of performance employed in this dissertation. OMB's Program Assessment Rating Tool (PART) scores, which serve as the dependent variable, are designed to measure how well agencies' programs achieve outcomes based on program purpose and design, strategic planning, and program management. Major studies of performance have used a similar definition. The Government Performance Project managed by The Pew Charitable Trusts and the Maxwell School of Citizenship and Public Affairs at Syracuse University, analyzes management capacity, which is the foundation for good results (Campbell Public Affairs Institute 1999, 2001). The focus on

effectiveness obviously excludes many other important aspects of performance, but effectiveness is the gold standard in this era of results-oriented government and it also dictates how the Office of Management and Budget (OMB) determines budget allocations.

> **Good government—a government responsible to the people whose dollars it takes to fund its operations—must have as its core purpose the achievement of results. No program, however worthy its goal and high-minded its name, is entitled to continue perpetually unless it can demonstrate it is actually effective in solving problems. In a results-oriented government, the burden of proof rests on each federal program and its advocates to prove that the program is getting results. The burden does not rest with the taxpayer or the reformers who believe the money could be better spent elsewhere (OMB 2004).**

This section of the chapter offers a brief review of the different methods of measuring federal government performance and, more importantly, maps the evolution of PART, the dependent variable used in the empirical analyses that follow. PART is by no means the first attempt to measure federal agency and program performance nor will it be the last. Below I offer an explanation for PART's development, analyze its pros and cons and how it is similar and different from previous efforts, and begin to offer some preliminary explanations for which factors might drive agency program performance at least during the "PART era."

There are many reasons to develop valid and reliable measures of government performance. Certainly one of the most important reasons to measure government performance is to reassure the public that the government is doing a good job with their tax money. Second, trust in the federal government declined dramatically in the second half of the twentieth century (Gallup 2008; Gore 1994) and there is a growing body of literature that suggests trust and social capital affect the performance potential of government (Tavits 2006; Knack 2002; Rice 2001; Putnam 1993). In addition to efforts

to restore public trust in government, measuring government performance is important because the size and nature of government responsibilities have changed dramatically in recent years. Theories about how best to manage government have changed accordingly; no longer are hierarchies necessarily viewed as the most effective way to delegate tasks and assign responsibility. Beginning in the late 20th century, management theorists and public administration scholars began to argue that hierarchies might actually impede production. More attention is being given to other administrative forms of governance such as horizontal, hybridized and associative governance. (Hill and Lynn, Jr. 2005)

The information technology revolution is a second key change in how theorists view government and how best to measure its performance. Information technology advances have made "flatter organizations" possible because individuals can take on wider ranges of responsibility and can communicate exponentially faster (Gore 1994). This means that the "top" and "bottom" of an organization can communicate more often and effectively and organizations can communicate directly with their constituents (Towns 2009). According to Beth Noveck, President Obama's deputy director for open government in the U.S. Office of Science and Technology Policy (OSTP), the administration is using tools like blogs to solicit public input on a range of issues. Currently the OSTP is using a blog to ask constituents how best to ensure scientific integrity in executive branch decision-making (Towns 2009).

Why do these changes matter for understanding and measuring government performance? When hierarchies were considered the ideal way to structure government agencies and organizations, it was believed that the layers of control over every input would provide appropriate checks and balances (Light 1995). The Johnson

administration's Planning, Programming and Budgeting and the Carter administration's Zero-Based Budgeting efforts for example, used administrative rules to guide tasks and budget allocations (inputs). Government performance was based on the relationship of costs and inputs or the idea that performance could be measured by cost reduction and improvements in operating procedures (Heinrich 2002). Now the way that government performance and accountability are determined by carefully measured results – outcomes and outputs (Gore 1994). President Johnson's Planning, Programming, and Budgeting System was the first serious effort to link program results with budget decisions. President Nixon's Management by Objective program sought to identify the goals of each agency program to limit redundant or ineffective programs. President Carter introduced zero-based budgeting in an effort to force government programs to prove their value every year. President Clinton's administration was known for its reinventing government efforts (OMB 2004). Most of these efforts made temporary improvements, although none had a lasting impact on "measurable accountability" with the qualified exception of the Government Performance and Results Act of 1993 (Radin 1999).

Beginning in the early 1980s and brought about by key changes implemented by the Reagan administration, government agencies at all levels began developing performance monitoring systems of various kinds (Wholey and Hatry 1992). The Government Accounting Standards Board (GASB) encouraged state and local governments to report indicators of service quality and outcomes in addition to financial data. Examples of outcome-oriented indicators at the state and local levels are evident in measures of school performance like student and teacher attendance rates and student achievement test scores.

The Reagan administration's New Federalism shifted management of many public programs from the federal to the state and local levels of government. A primary example of such a program is the Job Training Partnership Act (JTPA) established in 1982. A $5 billion program, management of the JTPA was highly decentralized and involved the private sector in local administration. Since the involvement of private industry and the decentralized management structure raised concerns about accountability, the Reagan administration developed new standards of performance management (Heinrich 2002). Barnow (2000) claims these new management standards were something completely novel and helped revolutionize government performance measurement and monitoring in three ways. First, they changed program measures from inputs to outcomes. While in previous years results of the JTPA would have been measured by inputs like the number of people trained, the new approach to performance measured results by the number of JTPA participants placed in jobs. A second change in the approach to performance measurement is the interlinking of measures across multiple levels of government. Third, the new approach provided incentives like budget increases for managers whose programs reached or exceeded established outcomes (Heinrich 2002).

The 1980s also brought about tremendous advances in statistical techniques used to measure government performance including experimental and quantitative evaluation. By 1990, the federal government began answering calls for a new focus on service quality and program results at the federal government level. The Chief Financial Officers Act of 1990 required federal agencies to provide systematic measurement of performance in addition to cost and financial data. Senator William Roth's "Federal Program

Performance Standards and Goals Act of 1991," asked federal agencies to develop plans to establish performance indicators to monitor outputs and results of major expenditures and to submit annual performance reviews to the President and Congress (Wholey and Hatry 1992). Also in 1991, the National Academy of Public Administration passed a resolution calling for government agencies to develop performance monitoring systems to report the program quality and outcomes. The resolution also suggested public participation in setting goals and monitoring results, using performance information to make subsequent program improvements, and establishing performance targets. As Frederickson and Frederickson (2007) conclude, "the scientific management-like reform initiatives in the 1990s emerged from the executive branch both as serious attempts to make government more effective and as a form of presidential politics."

Using Performance Measurement in Decision-Making

The developments of the 1980s and 1990s paved the way for the Government Performance and Results Act (GPRA) of 1993. Enacted with bipartisan support, the Government Performance and Results Act (GPRA) provided agencies with a roadmap to improve performance over the course of several years. Agencies were tasked with developing and implementing strategic plans, performance measures and reports to be used by congress and the executive branch to make budgetary decisions (Radin 1999). GPRA continued the trend of moving performance measurement away from process evaluations towards results and outcomes measures. Specific requirements of GPRA include a strategic plan that specifies agency goals and how they will be achieved; an annual performance plan that includes quantifiably-measured goals and performance

indicators and a specific level of performance to be achieved, and an annual performance report that compares actual and target performance goals (Heinrich 2002).

Most federal agencies developed their plans and performance measures on schedule, but few quantified their progress toward achieving key outcome measures. In 1999 and 2000, the Government Accountability Office (GAO) and the Senate Appropriations Committee reviewed agency compliance with GPRA. The reviews found that few agencies had made significant improvements, especially in measuring and achieving key outcomes. A majority of agencies were unable to produce credible data (Dull 2006). An example can be found in the GAO's 2000 report on the U.S. Department of Agriculture (USDA). Discussing the USDA's progress towards improved outcomes, the GAO found that the USDA:

> "uses a questionable methodology for measuring the success of its efforts to expand and maintain global markets for U.S. agricultural products. USDA's goals and indicators emphasize growth in the U.S. share of the global agricultural market—measured by changes in the dollar value of exports resulting from the implementation of trade agreements, market access enhancements, sales from annual trade shows, and agricultural exports. Yet, the dollar value of exports is subject to powerful external variables that transcend USDA's authority and ability to affect change in international trade. These variables include exchange rates, government policies, global and national economic conditions, climatic changes, and numerous other factors over which USDA has no control or strategies to address." (GAO 2000)

Instead of encouraging agency administrators to focus on results and performance, GPRA unintentionally shifted their focus to compliance with its procedural requirements (Gilmour and Lewis 2005). Agencies produce over 13,000 pages of performance plans annually, but they are rarely taken into account in the budgetary process (OMB 2004).

The use of performance measurement and information at the state level, like at the federal level, goes beyond budget allocation decisions. Performance information is sometimes linked to strategic planning, although with varying degrees of success. Performance measurement systems are well integrated into state governments having survived numerous changes in leadership. Performance information is used most commonly for program assessments, strategic planning and performance reporting. Most states have not begun to apply the information in targeted ways for personnel decisions, benchmarking across agencies or states, or for holding government officials accountable for their performance (or lack thereof) (Melkers and Willoughby 1998).

Melkers and Willloughby's 1998 study of the pervasiveness of performance-based budgeting in states found 31 states had legal requirements, 16 had administrative requirements, 1 had an executive order and 3 had no requirement. A majority of the states' legal and administrative requirements for performance-based budgeting were less than 5 years old and while most states required performance information to be generated and published, there was no specific direction for how program results, performance measures and state funding fit together.

In 2004, Melkers and Willloughby reanalyzed state performance-related requirements and found that all 50 states had either administrative or legal requirements for performance-based budgeting. Thirty-three states had maintained, amended or added to their requirements.

State requirements, both legislative and administrative, do not differ much in purpose from GPRA at the federal level. States have "sought to focus attention on outcomes rather than outputs and to link budget and managerial decision processes"

(Melkers and Willloughby 2004). Also like GPRA, many state performance requirements link performance to the strategic planning process. GPRA, however, is still more comprehensive than most state requirements because it also involves oversight of management, benchmarking and results.

States, like federal agencies, are "graded" on how well they perform. For the past ten years, Governing Magazine and the Pew Center on the States have collaborated to produce the Government Performance Project (GPP), which measures the quality of state management and performance. Like similar efforts to improve performance at the federal level, the focus is on program results. States, like the federal government, grade performance based on outcomes, not outputs. The grades are calculated based on extensive, ongoing research into four key areas of government management: information, people, money, and infrastructure.

Performance evaluation and measurement are included in the information category. States are assessed based on how well they produce results-oriented information or outcomes, rather than a detailed accounting of how much work they have done (Governing Magazine 2008). In 1998, the GPP found that few states produced performance-related information. Ten years later however, four out of five states received passing scores for their management of performance efforts. The GPP, like Melkers and Willloughby (2004), found that while states have made significant headway in producing performance data, there is still disconnect between performance information and its use in the budgeting process. One positive step towards the link between performance and budgeting is compliance by all 50 states with Generally Accepted Accounting Principles (GAAP), uniform standards for bookkeeping.

State and federal performance efforts are mirrored at the local level. Managing

for results (MFR) occurs at the municipal level too, although with more limited success

than at the federal and state levels. Recent research, however, indicates that the

implementation of MFR at the local level is on the rise because of the involvement of

several key organizations including the International City/County Management

Association (ICMA) and The Urban Institute (Harsell and Jones 2002).

While performance measurement efforts are well underway at all levels of

government, there are many challenges for government officials, especially in developing

appropriate measures of performance. The next section of this chapter addresses many of

those.

The Challenges of Performance Measurement

Policy makers and managers have to agree on the appropriate indicators of

program/agency performance (Wholey and Hatry 1992). Performance measurement is

complex and agencies have to consider a number of obstacles. First, reported outcomes

are the result of contextual factors, both agency-related and external conditions. External

factors include general economic and political conditions, while agency-related factors

include organizational climate and complexity and competition among agency programs

or functional units (Heinrich 2002). Distinguishing between agency and external

contributions to performance can be tricky. Performance measurement is not necessarily

meant to establish causality (did the agency or some outside event cause the result), but

results indicate to managers and the president and congress if an agency or program

merits a closer look. Wholey and Hatry (1992) provide the analogy of a sports team

manager who uses the game score to decide how to alter his play strategy.

A second obstacle is developing valid and reliable measures of performance. For some programs, developing appropriate performance measures is extremely difficult because of the nature of the program's work. According to Wilson (1989), agencies can be considered one of four types depending on how easily or not their activities and results can be observed and measured. The four different types of agencies are production, craft, procedural and coping organizations. "Production" agencies are those whose activities and outcomes are readily observable. Such agencies include the Social Security Administration and Federal Emergency Management Agency (FEMA). Service programs like those conducted by the Social Security Administration are measured in reliable and valid ways and produce short and long term measures of results like accuracy and response time for clients to receive checks. Production agencies are easier to manage because their activities and results are easy to oversee (Behn 2006). Wilson says production agencies are designed with "a compliance system to produce and efficient outcome" (Wilson 1989).

Craft agencies, in contrast to production agencies, are those with easily observed results, but whose activities are more difficult to observe. The armed forces are craft agencies, especially in wartime. Their activities are not easily observed, but the outcome of those activities is readily seen. In procedural organizations, activities and processes are observed, but the outcome remains hard to see. In peacetime, the armed forces become procedural organizations because their activities and preparation is observed, but the result or outcome of those activities are not apparent until needed. Coping organizations are the most difficult to measure and to manage because neither their activities nor results are observable. The best examples of coping organizations are

police departments and the Department of Education. It is very difficult to observe teachers conducting classroom lessons on a regular basis and it is hard to know whether and how much students are learning (Wilson 1989). Depending on which type of agency then, the management style and measurement of results will change. I will revisit this subject in Chapter 4.

A third obstacle to consider is what makes an "acceptable" performance level? For some programs, the ideal is perfection. In other words, the desired outcome is zero crimes committed, zero illness outbreaks and the like. For other programs, it is not as straightforward. Instead, programs must demonstrate a positive change in performance, compare actual to target outcomes, and compare performance against similar agencies and programs.

An additional obstacle requires agency program personnel to be ready and able to respond to the public and other branches about negative performance. Performance outcome measures can be misunderstood and misused and it is therefore the job of agency staff to be prepared to interpret measures clearly and effectively. Agency explanations should always accompany results reports (Wholey and Hatry 1992).

I have reviewed previous efforts to measure performance at various levels of government and some of the specific challenges associated with such efforts. Next I provide an overview of the Bush administration's program to measure federal government performance, which also introduces the dependent variable for the analyses that follow.

The Development of the Program Assessment Rating Tool

President Bush took office in 2001 with an agenda of improving government

management. Although GPRA had only begun a few years prior to his taking office,

President Bush took a different approach to performance management (Frederickson and

Frederickson 2007). A major component involved making government agencies more

results oriented by incorporating performance information into budgeting decisions. The

President's Management Agenda (PMA) was introduced in 2002. The PMA was made

up of five objectives: 1) improving strategic management of human resources, 2)

increasing the use of competitive sourcing in service delivery, 3) using appropriate

financial reporting tools to control agency resources, 4) improving the use of electronic

government tools, and 5) integrating performance results with funding decisions

(Greitens and Joaquin 2009; Breul 2007). Agencies were given a red, yellow or green

rating for how well they met each objective with green being the optimal rating. During

the seven years the scorecard was in effect, the number of agencies achieving "green"

scores increased from 15 to 85 percent (Barr 2008).

While the increase in scores over time is some evidence that the OMB scorecard

was successful, it was met with strong resistance by agencies and congress because of the

controversial objective of increasing outsourcing of government jobs. Congress

eventually limited outsourcing of federal jobs, which remained a point of contention for

the remainder of Bush's second term. The final OMB scorecard is found in Table 1.1.

[Insert TABLE 1.1]

The fifth objective of the OMB Scorecard Initiative called for agency funding

decisions to be linked to performance, which was one of the intended, but unfulfilled

goals of GPRA. While GPRA helped to refocus agencies on results by requiring annual

goal statements, strategic plans and both output and outcome measures, a GAO analysis

of GPRA found that this information, while useful, was not utilized by policymakers

(Gilmour 2006). President Bush first directed the OMB to help agencies incorporate

performance information into budget requests. OMB with the help of agencies and

outside experts then developed an assessment framework called the Program Assessment

Rating Tool (PART). Bush developed PART in addition to GPRA for a couple of

reasons. First, managing the government "better" than one's predecessor is a main theme

of any presidential campaign. Second, GPRA was not meeting the expectations of

performance management experts (Frederickson and Frederickson 2007). The intention

of PART was to hold failing agencies and programs accountable through budget

reductions or elimination and GPRA lacked such accountability. The PART grading

system was heralded by the administration as a rich new data source for evaluating

agency performance.[1]

Several groups participated in the development of PART. Marcus Peacock,

former associate director for natural resources, energy and science at the Office of

Management and Budget and Deputy Secretary of the EPA, is credited as the "father of

PART." Peacock led the effort to develop PART at the request of then-President Bush

(Peacock 2008). The first draft of the PART questionnaire was circulated for public

comment in May 2002 and then reviewed by a number of groups including the

Performance Measurement Advisory Council and the President's Council on Integrity

and Efficiency. The first group was chaired by Mortimer Downey, former Deputy

---

[1] The Obama administration is currently conducting a major review and overhaul of PART. I will discuss
these efforts in detail in chapter 5.

Secretary of Transportation (OMB 2004). Subsequently, the National Academy of Public

Administration convened a special workshop to review PART. A congressional hearing

was also held before the final version was approved July 10, 2002.

PART differs from previous federal performance management efforts in a few

ways. First, PART was designed to be an integral component in the development of the

executive budget and President Bush promised to use PART to target program

reductions. The evidence that this actually occurred, however, is mixed. In the

president's FY 2005 budget, for example, 65 programs were recommended for

elimination. Thirteen of the 65 programs were recommended based on their PART score

and only one of those 13 was actually eliminated (Dull 2006).

Second, there is evidence that programs receiving scores of adequate or better are

generally recommended for budget increases, while programs that receive ineffective

ratings are more likely to be recommended for budget cuts (Frisco and Stalebrink 2008).

A third difference between PART and previous performance initiatives is the

amount of resources committed to it. The President, OMB staff, and agency resources

were devoted to completing PART.

How PART Works

The Program Assessment Rating Tool, or PART, was part of the Bush

administration's broader Performance Improvement Initiative (PII) and it fulfilled the

executive order E.O. 13540 "Improving Government Program Performance." The PII

institutionalized the policy objectives of EO13540 and had two key goals: first to

improve program performance, second to make a greater investment in successful

programs.

PART is a 25-part questionnaire made up of mostly "yes" or "no" responses with the exception of one section of answers "small extent" or "large extent." Agencies worked in tandem with OMB staff to complete the questionnaire and to determine whether the evidence agencies provided for their answers was sufficient. Each program received a rating of 0-100 on each of four parts (Program Purpose and Design, Strategic Planning, Program Management, and Program Results) which were weighted and totaled, receiving an overall qualitative rating of effective, moderately effective, adequate, ineffective, or results not demonstrated (RND). It is important to note that regardless of the overall score, programs received a score of RND if OMB determined that performance measures were unacceptable or that performance data were not collected to determine whether the program was actually performing (ExpectMore.gov). In Chapter 3 I take a closer look at other explanations for why programs were given the RND rating.

Programs that received one of the first three ratings were considered to be "performing," while programs rated ineffective or RND were "not performing." With their overall rating, agency programs also received key findings from the evaluation as well as an improvement plan developed in collaboration with OMB. All of this information remains available on ExpectMore.gov. By the end of the Bush administration, more than 1000, or 98 percent of all federal agency programs, were evaluated via PART (For question categories and sample questions, see Table 1.2).

[Insert Table 1.2]

Agencies worked with OMB staff to determine which programs were to be reviewed each year. Once programs were selected for assessment, agency personnel were responsible for completing the PART survey, although they did so in close

communication with OMB staff. Personnel appointed to this task included day-to-day

program staff, budget staff who handled resource requests and justifications, and

planning and performance staff who monitored program results and impact. OMB

assigned a program examiner and also solicited input from staff at OIRA and the Office

of Federal Financial Management (Gilmour 2006).

Agency staff completed the PART questionnaire via PARTweb, an online

application. PARTweb allowed for ongoing virtual collaboration with the designated

OMB examiner. PART was completed according to a detailed schedule published at the

beginning of each year that was a joint determination of agencies and OMB.

Prior to completing the PART survey, program staff did four things. First, the

program or "unit of analysis" for PART was determined at the agency level. What

constituted a "program" varied from agency to agency. Agencies usually referred to the

budget structure to determine program boundaries. While a program was most often

thought of as a single program, a program could also be a collection of several programs

managed as a single unit, such as the Administration on Aging's grant programs (PART

2008).

The second step was for agencies to designate the program type among seven

different categories. This was important because there were specific questions in PART

for each type of program as well as common questions across all program types.

Programs received mixed assessments when they employed more than one "mechanism"

or type to achieve their goals. Question weighting was an important third step. The

relative weights of the questions were adjusted prior to completing PART. Weights

emphasized key factors, especially when a program's funding was determined by policy and not performance (OMB 2008).

The final and most difficult step was to select and apply performance measures; I discussed many of those challenges in Chapter 1. Performance measures were developed many years prior under GPRA and were supposed to be the measures used for PART. However, there were many disagreements between OMB and agency staff as to whether these measures were "appropriate," which echoes some of the tensions between GPRA and PART that I discussed in a previous section of the chapter.   Again, all programs were not measured the same. There had to be a clear justification and rationale for whichever type of performance measure used and non-experts had to be able to understand what was being measured and how it was relevant.

There are important distinctions among the different types of measures. Outcomes are the preferred measures and are defined as the intended result of carrying out the program. An outcome is an external event or condition that is of importance to the intended beneficiaries of a program. An outcome could be the reduction in new HIV cases in the U.S. Outputs measure the level of activity provided over a period of time and relate to the internal activities of the program. The percent of warnings more than 20 minutes prior to a tornado is an example of an output measure. Outputs support and lead to outcomes. Outcomes are given more importance because they are the ultimate results of a program (Gilmour 2006).

Efficiency measures are also significant because they determine the ratio of benefits derived from a program to resources utilized. The Administration's PII required efficiency measures to achieve "green" status. An outcome of an efficiency measure

captures the improvement in program outcomes for a given level of resources expended. An output of an efficiency measure demonstrates how to produce a given output level with less resources. For example, the ratio of the number of benefit checks mailed to the amount of money spent on labor.

Target and baseline measures are also important to demonstrate program improvement. The baseline is the starting point from which gains are measured and targets are established. A target is the improved level of performance required to achieve stated goals. Programs have long-term and annual performance goals. Quite often the two are combined so that progress toward a long-term goal is measured on an annual basis.

The PART survey was made up of four sections and each was weighted according to its importance in driving results. Section One was Purpose and Design and asked whether they were clear and sound (e.g., "Does the program address a specific and existing problem, interest or need?"). Section Two was Strategic Planning and assessed the validity of the long-term and annual performance measures and targets (e.g., "Does the program have ambitious targets and timeframes for its long-term measures?"). Section Three evaluated the management of the program, financial oversight and improvement efforts (e.g., "Does the program collaborate and coordinate effectively with related programs?"). The final section, results and accountability, rates program performance on measures and targets reviewed in the strategic planning section (e.g., "Has the program demonstrated adequate progress in achieving its long-term performance goals?"). The Purpose and Design and Management sections each made up 20 percent of the rating. Strategic Planning accounted for 10 percent while Results

accounted for 50 percent, or half of the score. Most of the questions were answered either "yes" or "no." While it sounds simple, there was a high level of evidence required to answer "yes" because it indicated satisfaction with all requirements of a given question.

Once a program received a quantitative score according to the total of the weighted scores, it received one of the previously explained qualitative ratings. The numerical range of each rating is found in Table 1.3.

[Insert Table 1.3]

Agencies did not always agree with OMB, so they could formally appeal their PART score. Agencies could request a reassessment, but had to meet several requirements. They had to demonstrate that a program had made significant efforts to improve performance. Appropriate evidence might include new outcome data or performance measures. PART included questions about steps programs had undertaken to improve performance. Agencies had to provide draft text for the explanation and evidence and a link to or copy of the relevant evidence. The appeals were judged by a five-person panel made up of members of the President's Management Council. The Council was comprised of deputy secretaries responsible for management improvement from various agencies (ExpectMore.gov). Programs rated RND received the most pressure to follow OMB's recommendations for improvement and to complete another PART assessment. This determination involved some subjectivity on behalf of the OMB. For this reason, I will perform a selection model analysis of programs rated RND in Chapter 3.

After the PART score and rating were determined, program staff worked with OMB to develop an improvement plan that outlined management actions the agency would take to improve performance. Most improvement plans included action items to be completed within a year. An agency's progress toward meeting their improvement plan was updated twice yearly on ExpectMore.gov. Making this information available via ExpectMore.gov represented a sea change in how performance management information was previously handled because it increased the access and transparency of the PART process.

PART was labor intensive and required a great deal of organization and dedication on behalf of the program staff. According to Gilmour (2006), programs faced four key challenges to meet the demands of PART. First, departments and agencies had to organize appropriately. Second, the PART questionnaire was the main vehicle for articulating a program's accomplishments and shortcomings, so it had to be done carefully. Third, and perhaps most importantly, programs had to develop suitable performance measures. As will be explored below, this was more difficult for some types of programs as their outputs/outcomes were not easily observed. Without measures that are agreeable to OMB, programs risked receiving the results not demonstrated rating. Fourth, programs determined the extent to which program managers were responsible for performance measures and results (Gilmour 2006).

Criticisms of PART

While PART was arguably the most comprehensive performance management program implemented at the federal level, it was not without its shortcomings. I acknowledge at the outset that these are imperfect data. First, unlike GPRA, PART

focused exclusively on executive priorities, which can be to the detriment of other stakeholders and constituencies. Second, the objectivity of PART graders was frequently called into question. There were a number of steps in the process where subjective judgments were made by OMB. For example, OMB determined what constituted "acceptable" performance measures and "adequate" data, although criteria for those determinations was elusive. Third, programs selected one goal, while many had multiple and competing goals. I will explore goals and performance further in Chapter 4. Fourth, some argued that there was a bias against certain program types (e.g., research and development) because there were no suitable outcome measures, the cornerstone of how PART evaluates results. Fifth, PART, like GPRA, required many hours of OMB and agency personnel's time (Frisco and Stalebrink 2008). With the exception of how labor intensive PART was, I address these criticisms in the quantitative (chapter 3) and quantitative (chapter 4) analyses. I control for different types of programs to determine whether there are systematic differences in PART scores according to program type. I examine programs rated RND in a separate analysis to uncover whether the rating is attributable to political factors or bias. I also provide a measure of agency ideology to determine whether programs housed in "liberal" agencies have more difficulty scoring well on PART.

From its inception, PART was heralded as the "new and improved" performance management program that was to rectify some of GPRA's shortcomings. GPRA is a statute that provides the framework under which agencies prepare strategic plans, performance budgets, goals and report on how well they achieve those goals. PART is a "systematic method of assessing performance of program activities, focusing on their

contribution to an agency's achievement of its strategic and program performance goals"
(Frederickson and Frederickson 2007).

PART was to give GPRA 'teeth' as it was supposed to influence the president and
congress' budget decisions (Gilmour 2006). According to a recent analysis however, the
link between GPRA and PART appears questionable at best. Gueorguieva, et. al. (2008)
analyzed the response of seven agency programs to both GPRA and PART and
concluded the link between the two programs is relatively weak. The two programs
differ in five key areas: focus, government branch involvement in the process,
organizational approach, requirements, and approach to performance measures
(Gueorguieva et al. 2008). The differences are illustrated in the Table 1.4.

[Insert Table 1.4]

The differences between PART and GPRA outlined in Table 1.4 are not
insignificant. Federal agency programs vary, as has been discussed. PART required
agency programs to develop outcome measures and rewarded programs that did so. As
mentioned above, some programs may have outputs and outcomes that are difficult to
observe. GPRA allows for variance among program types by allowing output measures
while encouraging development of outcome measures over time. PART, by contrast,
required programs regardless of type to develop and employ outcome measures, although
proxies were acceptable in certain cases. In Gueorguieva et al.'s (2008) analysis of seven
programs' scores on GPRA and PART, they found that the same program scored quite
differently.

According to OMB, GPRA and PART serve different purposes. GPRA provides
a framework for agencies to prepare strategic and performance plans, and reports that set

goals and benchmarks for reaching those goals. PART was a systematic program for assessing performance based on agencies' achievement of strategic and performance goals. Ideally GPRA and PART worked together to help agencies develop an infrastructure to improve performance over time. There was only partial overlap, however, in the performance criteria used between GPRA and PART to evaluate programs. According to Gueorguieva et al. (2008), there are several specific points of conflict between these two performance initiatives.

First, while GPRA and PART emphasized outcome measures, there was no dialogue between the two programs about which measures were used for evaluation. Programs scored quite differently on PART versus GPRA because PART did not always consider previously-reported GPRA performance measures. Evaluations between PART and GPRA then most likely differ because they used different performance criteria. While outcome measures are encouraged under GPRA, there was more emphasis on end outcome measures under PART. According to Gilmour (2006) who conducted interviews with agency staff responsible for PART, agencies whose programs received too many RND (results not demonstrated) scores risk receiving a red on the president's management agenda or OMB scorecard. As I mentioned earlier, the OMB scorecard rates how well major departments and agencies are executing the five major initiatives: Improved Program Performance; Strategic Management of Human Capital; Expanded Electronic Government; Improved Financial Performance; and Competitive Sourcing (Results.gov 2008). There are corresponding "standards of success" for each of the three ratings: green, yellow and red. The improved program performance initiative requires that agency to employ at least one outcome measure per strategic goal and program to receive a "green" rating.

Second, GPRA and PART present conflicting information about program performance. Even though different programs might meet an equal number of performance goals under GPRA, they receive different PART scores. A third area of conflict between PART and GPRA relates to PART's use of criteria that either conflict with or are peripheral to a program's statute or intent. Some programs are designed to have broad goals. This is true for federal grant programs. Their goals are broad so states that receive monies can define goals more narrowly based on their unique needs. Fourth, PART did not consider the congressional intent behind a program's original statutes (Gueorguieva et al 2008).

A fifth point of conflict between PART and GPRA is the level of analysis. This might be the most crucial point of departure between these two programs. GPRA asks agencies to develop mission statements, goals and objectives, as well as descriptions of program evaluations. PART only looked at the program level (Radin 2006). Sixth, PART asked programs if they duplicated the efforts of related programs. If so, the program was penalized. However, PART also asked programs the degree to which they collaborated and cooperated with related programs. This placed a contradictory emphasis on related programs according to Gueorguieva et al. 2008.

A final difference between PART and GPRA is the extent to which external factors are accounted for in performance evaluations. GPRA asks agencies to identify external factors both within and outside of the agency's control that could impact performance. The only PART question to evaluate external factors was about how much partner performance affected program results, which it assumed was within the control of

the agency. "Partners" refers to a wide range of governmental and non-governmental actors, such as state and local authorities and independent contractors.

I concur with many of the critics that PART is an imperfect system, but it does provide an unprecedented amount of data with which to analyze more than 1,000 agency programs that will hopefully shed some light on a basic, but important question: why some agencies (or programs) perform better than others. This dissertation is not a referendum on PART. It does not answer the question of how well PART evaluates agency performance; rather it is about what factors help explain agency performance and PART serves as a proxy measure of performance. Alternative measures include surveys of federal government staff as Lewis (2008) and others have used as proxies for performance. These surveys do tap perceptions of performance, but they are also flawed. I will revisit PART's limitations in my discussion of the empirical results. I now look at the extent to which PART has impacted performance management in the federal government.

PART's Impact

The end of 2008 marked the conclusion of the first five-year PART evaluation cycle. More than 1,000 agency programs were scored. Since PART's inception, questions have been raised both inside and outside of government as to what impact PART has and will have for how agencies and their programs perform. Beyond performance, there is also the question of how the performance information gathered would be used by Congress, which has to approve agency and program budgets. How aware was Congress about PART? Did members of Congress consider PART scores when making budget

decisions? To what extent were well-performing agencies rewarded and poor-performing agencies punished via the appropriations process? (Frisco and Stalebrink 2008).

Table 1.5 shows the number of agency programs per rating category. The agencies are listed in order by how many effective ratings programs received minus those with RND scores. There is a great deal of variation among agencies included in Table 1.5. What might explain this variation? The difference in agency missions? Size? Budgets? Possible explanations of performance are explored in detail in the following chapters, but another important question is what happens to agencies, like the Department of Education, that have the highest percentage of poorly performing programs?

[Insert Table 1.5]

A limited number of scholars (Frisco and Stalebrink 2008; Radin 2006; Gilmour and Lewis 2006; Mullen 2006) have explored Congress' use of PART as well as its impact on improved performance. Frisco and Stalebrink (2008) analyzed congressional committee hearing reports for PART-related content and found that PART was mentioned in discussions about budget decisions. The relative influence of the PART score, however, could not be determined. Frisco and Stalebrink (2008) concluded that Congress learned about PART via external actors. OMB provided PART data as part of its budget justification and agencies included PART scores in congressional hearing testimony.

Researchers appear to be in agreement that PART scores did not have direct repercussions for agency program budgets or long-term survival. PART was an important tool that may have had some impact on budget and management decisions, but several other factors were taken into account.

Discussion

In this dissertation, I provide a theoretical and empirical challenge to the predominance of the principal-agent model as applied to the relationship between the president, congress and the federal bureaucracy. In this model, the president and congress are the "principals" who must delegate tasks to experts or "agents" within the bureaucracy. However, by delegating such tasks, the principals run the risk that agents will act in accordance with their own motivations commonly referred to as a moral hazard problem. To ensure compliance, presidents and congress use external control mechanisms such as agency design and budgeting to influence agency behavior. As discussed earlier, some studies suggest this external control could negatively influence performance, although the empirical evidence is mixed.

Since the external control theory cannot adequately explain performance in empirical models, scholars have recently begun to question the predominance of the principal-agent model as applied to the president-congress-bureaucracy relationship. A major weakness of the theory is its failure to recognize that no true "top-down" command structure can effectively operate in our separation-of-powers system of government. Certainly controls exist, but they do not appear to have the intended effect of influencing agency behavior, at least not all of the time.

The primary weakness of the external control theory, however, just may be the failure to include the bureaucracy in the explanation of performance. Bureaucratic values are often not accounted for in explanations of performance (Meier and O'Toole 2006). Why might this matter? It could very well be the case that agencies perform according to internal phenomena like culture and leadership. Performing well builds their reputation,

which is important for working with other agencies and for attracting constituency support. Individuals within agencies are committed to "democratic norms" such as public service and are motivated to do their job well and help the agency perform well. As Dahl (1947, 1970) stated many years ago, 'The primary control on administrative behavior is the inner check, the values held by the bureaucrat.'

This dissertation advances a theory of agency performance, not by dismissing the external control theory altogether, but by completing the theoretical puzzle of agency performance by including internal agency characteristics.

Preview of Chapter 2

Chapter 2 introduces my approach for quantifying agency performance. I provide a thorough review of the data and methodology and I explain the set of variables selected to measure both the external and internal characteristics of agency performance. The variables used have been part of many previous analyses of agency performance, but not in the same combination.

TABLE 1.1	OMB President's Management Scorecard (As of July 2008)					
	Energy		**Transportation**		**Environment**	
	Status	Progress	Status	Progress	Status	Progress
Agriculture	Green	Green	Red	Green	Red	Green
Commerce	Green	Green	Red	Green	Yellow	Green
Defense	Green	Green	Red	Green	Red	Yellow
Education					Red	Yellow
Energy	Green	Yellow	Yellow	Green	Yellow	Green
EPA	Green	Green	Red	Green	Yellow	Green
HHS	Red	Green	Red	Yellow	Yellow	Green
DHS	Yellow	Green	Red	Yellow	Yellow	Green
HUD	Green	Green	Red	Green	Yellow	Green
Interior	Green	Green	Yellow	Green	Red	Yellow
Justice	Red	Green	Red	Green	Red	Green
Labor	Yellow	Green	Green	Green	Yellow	Green
State	Red	Green	Red	Green	Red	Yellow
Transportation	Green	Yellow	Red	Green	Red	Red
Treasury	Green	Green	Green	Green	Yellow	Green
Veteran's Affairs	Green	Green	Red	Green	Yellow	Green
Court Services	Green	Green	Red	Yellow		
GSA	Green	Green	Red	Yellow	Red	Yellow
NARA	Green	Green			Red	Red
NASA	Green	Green	Red	Green	Yellow	Green
OPM					Yellow	Green
SSA		Green	Yellow	Green	Red	Red
Smithsonian			Red	Green	Red	Red
TVA	Green	Green		Green	Yellow	Green
USPS	Red	Green	Red	Green	Red	

Key for TABLE 1.1

- **Green** for success,
- **Yellow** for mixed results, and
- **Red** for unsatisfactory.

TABLE 1.2	Description of PART Sections and Questions		
Question Category	**Description**	**Weight (Percent of Total Score)**	**Sample Questions**
Program Purpose and Design	Rates whether the program purpose and design are clear and defensible	20%	Does the program address an existing problem, interest or need? Is the program designed so that it is not redundant or duplicative of any other federal, state and local or private effort?
Strategic Planning	Rates if the agency sets valid annual and long-term goals for the program	10%	Does the program have ambitious targets and timeframes for its long-term measures?
Program Management	Rates agency management of the program including financial oversight and program improvement efforts	20%	Does the program collaborate and coordinate effectively with related programs?
Program Results	Rates program performance on goals outlined in strategic planning section and through other evaluations	50%	Does the program demonstrate improved efficiencies or cost effectiveness in achieving program goals each year?

Source: Gilmour 2006

TABLE 1.3	PART Rating Categories
Qualitative Rating	**Numerical Score Range**
Effective	85-100
Moderately Effective	70-84
Adequate	50-69

Ineffective	0-49

TABLE 1.4   **GPRA v. PART: Overview of Key Differences**

Issue:	GPRA	PART
Focus	Offices and Organizational Units	Programs
Branch Involved in Process	Legislative and Executive Branches	Executive Branch only
Organizational Approach	Bottom-up; starts with program units	Top-down; starts with OMB approval of measures
Requirements	Multiple: strategic plan, performance report	Performance measures
Approach to Measures	Multiple types – outcomes are highlighted	Efficiency outcomes

Source: Gueorguieva et al. 2008.

TABLE 1.5   **Agencies Ranked by the Percentage of Programs Rated Effective Minus Results not Demonstrated**

	Percentage Rated Effective	Percentage Rated Results not Demonstrated (RND)	Effective Minus RND
State	38	7	31
Treasury	44	16	28
Defense	34	13	21
Transportation	20	0	20
Energy	22	8	14
Commerce	18	18	0
Labor	4	11	-7
Justice	11	19	-8
Environmental Protection Agency	0	13	-13
Health and Human Services	11	27	-16
Agriculture	6	27	-21
Homeland Security	16	38	-22
Housing and Urban Development	4	32	-28
Interior	8	37	-29
Veterans Affairs	0	33	-33
Education	3	55	-52

CHAPTER TWO

External and Internal Determinants of Agency Performance: An Introduction to the Data

Introduction

The aim of this dissertation is to evaluate different explanations of agency performance. As detailed in chapter 1, principal-agency theory claims agencies are subject to external tools of political control by the president and congress, which drive behavior, outcomes and performance. Among such tools are agency structure, appointments and budgeting, although I acknowledge there are a variety of conclusions about how exactly these tools are used and whether they are successful. An alternative theory is that internal agency elements determine performance and that agencies can and do exercise political independence (Potoski 1999). According to this theory, there are many factors that mitigate external political control of agencies such as bureaucratic values and culture and strong leadership.

This dissertation attempts to build a holistic model of agency performance that includes variables representing both external and internal determinants of performance. Previous efforts, both within political science and public administration, have excluded independent variables which might have led to incomplete conclusions about what affects agency performance. My approach is to run external and internal-only models as well as combined models to determine how different variables affect performance. Each of these

variables, how they are measured, and my expectations for how each affects

performance, are explained below.

Chapter Preview:

In this chapter I review the literature and support for the alternative theories of

agency behavior and performance. I begin with what I refer to as external determinants

or characteristics that potentially affect agency performance. These include agency

politicization as measured by the ratio of appointed to career staff within an agency, and

program budget size. Internal determinants include ideology, leadership, constituency

support and program type. Within each section I review the literature and state my

hypotheses for how I expect each variable to affect agency program performance.

External Determinants of Performance: Divided Government, Budget Size, Percent Appointed

Presidents, like congress, face tradeoffs when exercising control over the

bureaucracy. Congress attempts to control the bureaucracy through ex ante controls (e.g.,

legislation) and/or ex post oversight (e.g., hearings). The president uses unilateral action,

structure, budgetary controls and appointments to control agencies (Moe and Howell

1999; Lewis 2003). However, while the president, and congress, attempts to control

agencies are meant to yield more responsiveness, research shows rather that performance

can be negatively affected (Meier and Smith 1994). Additional control mechanisms

include monitoring of behavior and actions (Huber and Shipan 2000), providing

incentives (Heinrich 2002), enforcing sanctions (Calvert, McCubbins and Weingast),

designing agencies (Lewis 2003) to be responsive to the enacting coalition, and

administrative procedures. Each has been examined extensively within the political science literature.

Boyne finds in his 2003 analysis of more than 60 empirical studies that resources variables are consistently shown to be critical to improving agency performance. The most obvious and, arguably important, of these resource variables is the program's budget. In their 2006 article "Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores," Gilmour and Lewis find some evidence that programs with smaller budgets do not perform as well as those with larger budgets; therefore,

Hypothesis 1: As program budget increases, program performance will increase.

The politics at the time of agency creation is a popular explanation for how politicians control the bureaucracy. Throughout the 20th century, presidents increased their power over the executive branch by centralizing control over appointments in the White House, using appointments to change policy and by reorganizing agency bureaucracies. To increase their ability to control the bureaucracy, presidents such as Truman and Nixon have sought to reduce the number of agencies in the hope it would make their job easier (Lewis 2003). Presidents want agencies that are responsive and not insulated from their control. Congress also has a say in how agencies are designed and it might seem all too obvious that members would not want agencies to be responsive to the president. Congress, however, acts in its own best interest and realizes that agencies are responsive to the president often make for effective policy implementation in their districts, which helps their reelection chances. If Congress insulates agencies from presidential control, they run the risk of limiting the agency's effectiveness. Moe (1989)

argued that while presidents and sometimes Congress seek to design structures that make

agencies responsive to the president, or what he termed an agency that has "responsive

competence," it did not necessarily make agencies more effective. According to Lewis

(2003), when government is divided there is a greater probability that a new agency will

be insulated from presidential control if the majority in Congress is large enough to win a

battle with the president over insulation. If Lewis was right, then agencies created during

divided government will not perform as well as agencies created during unified

government because the partisan wrangling creates agencies with a layered, insulated

structure not conducive to quality performance (Lewis 2003). Therefore:

Hypothesis 2: If control of government is divided at the time of agency creation, then performance will be lower.

Agencies that are filled with a greater number of appointees tend not to perform as

well as those run by careerists, because appointees are often not as prepared or qualified

as careerists and serve far shorter tenures (Lewis 2008). David Lewis' book *The Politics*

of Presidential Appointments (2008) explores agency politicization. The president and

members of congress attempt to control the bureaucracy via politicization, which means

appointing staff who share their policy interests and those who are loyal supporters.

Lewis developed four key hypotheses about when and why agencies become

more politicized. First, presidents are more likely to politicize when their policy views

diverge with an agency. Second, politicization will decrease when it is likely to

negatively affect agency competence. In other words, politicization is less likely within

agencies with complex or technical tasks (Lewis 2008, Wilson 1989). This is the most

direct proposition for how appointments affect performance. Third, politicization

increases when the same part controls the White House and congress. Fourth,

politicization will vary by the type of appointment. Senate-appointed positions (PAS) have the most impact on policy change but are the most difficult to change under a watchful Congress. Schedule C appointments are easier to fill with campaign supporters and allies because they are less prestigious and less visible. Because of these differences, Lewis believes politicization among PAS positions will increase when the policy views of the president and an agency diverge. By contrast, politicization among Schedule C positions will increase in agencies with policy views similar to the president (Lewis 2008).

Using data from 1960-2004, Lewis finds that neither of the common perceptions about politicization hold true. Politicization has not increased over time and Republican presidents are not the primary drivers. In a subsequent econometric analysis, Lewis tests and finds support for his four hypotheses using data from the past three presidents: George H.W. Bush, Bill Clinton and George W. Bush. For the first hypothesis about policy disagreement with the president, Lewis finds that despite an initial decline in politicization after a party change in the White House, there is a long-term increase in politicization after such a change. There is greater increase of politicization in agencies with policy views dissimilar to the president. He also finds that when an agency has a large number of professional and technical staff, politicization is lower. Presidents, therefore, resist increasing appointments in agencies whose performance might suffer. Not surprisingly, Lewis finds that politicization decreases when the White House and congress are controlled by different parties. It is not in Congress' best interest to support politicization by a president of the other party. And, finally, patronage or Schedule C appointments increase among agencies with policy views similar to the president.

Lewis used PART scores from FY 2004-2006 to examine whether systematic differences in scores exist according to whether agency programs are managed by appointees versus careerists. Indeed the programs managed by appointees have significantly lower PART scores than those managed by careerists, which is especially true for the program management and results section. This is noteworthy because the results section counts for half of the overall score. When controls such as budget size, program type and ideology are added to the model, Lewis finds that the previous results hold.

Hypothesis 3: As the ratio of appointees to careerists increases, the overall performance score will decrease.

While the above literature reflects the negative influence of external factors on performance, recent quantitative examinations fail to show any influence on performance. Using agency structure as a key explanatory variable, Krause and Douglas (2005) found agency outputs were unaffected in an ex ante fashion by the extent to which agencies are insulated from political influence. The insulation hypothesis predicts agencies will perform differently if they are designed to be insulated from political pressures. Additionally, insulated agencies should produce more accurate and less biased policy information because of their greater capacity to resist political pressures. However, in a comparison of agencies performing similar tasks, Krause and Douglas found that insulated agencies perform no differently than non-insulated agencies. Their explanation is what they term the "reputational concern" of agencies. According to this explanation, agencies will not respond to political influence because of a stronger concern with their reputation among other agencies. This concern will lead agencies to behave similarly in an effort to gain autonomy and prestige for the quality of their outputs (Krause and

Douglas 2005). This evidence suggests that the internal workings of an agency may be important predictors of performance.

The mixed empirical evidence for the external explanations of performance, including the structural and insulated agency hypotheses, naturally leads to questions about what other factors explain performance variance among agencies. The political control theory is widely accepted, but the empirical evidence leaves room for other explanations (Meier and O'Toole 2006).

Internal Determinants of Performance: Agency Ideology, Leadership, Culture, Constituency Support

Moving beyond the political control theory, I am in search of alternative, perhaps complementary explanations of agency performance. Scholars have examined how internal agency culture and related characteristics affect performance. Agency ideology, leadership, and culture vary a great deal. Some agencies have a "strong sense of mission" and have a distinctive culture, while others are more politicized (Wilson 1989; Lewis 2008). Some agencies might be dedicated to fulfilling "democratic" norms and others less so (Meier and O'Toole 2006). All of these internal agency characteristics have the potential to impact performance.[2]

As part of the discussion of how agency characteristics affect performance, the nature of the program's work is particularly important. Report cards used to demonstrate performance trends and differences in performance across agencies in the 1999 Government Performance Project reveal several potential explanations for agency performance. Agency programs that distribute money, enjoy diffuse support, and are

---

[2] As I discuss in Chapter 3, one potential limitation of my empirical analysis is that my independent variables are not "independent," but rather functions of one another. For example, leadership likely shapes culture, which in turn influences ideology. I recognize this as a potential shortcoming.

production oriented perform well. By contrast, agencies whose primary task is to collect money have ambiguous or conflicting missions, and those subjected to coercive control perform poorly. The Internal Revenue Service is a prime example (Barrett and Greene 1999). What strong performing agencies appear to have in common are discretion and the ability to operate somewhat independently. The Social Security Administration and the U.S. Postal Service are production organizations with clear outputs and outcomes that make it easy for agency leaders to see accomplishments. These two agencies also enjoy diffuse support which improves performance. Support is important for agencies, so they are concerned about how their reputations are perceived by a number of different parties including the president and congress, but particularly their constituents. Depending on the type of agency, constituents are those that receive agency goods and services (distributive and redistributive agencies) and those whom the agencies are meant to serve and protect (regulatory). Agencies with greater constituency support have the means by which to withstand congressional and presidential interference (Lewis 2004; Riccucci 1995; Clarke and McCool 1985).

> *Hypothesis 4: Agencies with greater constituency support will have higher performance scores.*

Studies have also found that leadership affects performance. A 2004 GAO report on the status of performance-based management reforms states, "The clear and sustained commitment of an agency's top leadership to change is perhaps the single most important element of successful transformation and management improvement initiatives." Agencies with leaders who possess requisite experience and expertise, and who foster cooperation and espirit de corps among managers and operators tend to perform well. Boyne (2003) reviewed more than sixty empirical studies focused on improving agency

46

performance and concluded that internal resources and management are the most

important predictive variables. Specifically, leadership style and expertise are valuable

indicators of performance. Using Boyne's framework, Brewer (2005) developed a study

of frontline supervisors to better specify how management affects performance. Using

the U.S. Merit Systems Protection Board Merit Principles Survey (2000), he used federal

employees' responses to questions about their immediate supervisor's competence and

ability to motivate staff to determine the extent to which leadership can impact

performance. Brewer found that competent, skilled leaders that motivate employees are

important contributors to positive performance, therefore:

> *Hypothesis 5. If agency staff have positive perceptions about their agency's*
> *leadership, the agency will have a higher PART score.*

Agency culture also plays a vitally important role in how agencies perform and

this role has been somewhat overlooked in other examinations of performance. Many of

the explanations provided by the Government Performance Project (Barrett and Greene

1999) tap various aspects of culture: experienced leadership, intra-agency cooperation,

and clearly defined tasks. Organizational culture is broadly characterized as the

"persistent, patterned way of thinking about the tasks of the agency, how staff interact,

and how appropriate actions, rewards and consequences are defined" (Wilson 1989).

Wilson (1989) wrote nearly 20 years ago about the potential performance differences

between agencies with a single, well-defined culture characterized by a strong sense of

mission and those lacking such culture. He hypothesized that agencies with a single

culture will refuse to take on tasks delegated by congress and the president if they were at

odds with their "central mission" because it would upset the longstanding administrative

culture. March and Olsen (1989) echo this assertion by claiming "behavior is often

constrained or dictated by such cultural dicta and social norms. Action is often based more on identifying the normatively appropriate behavior than on calculating the return expected from alternative choices. Routines are independent of the individual actors who execute them and are capable of surviving considerable turnover in individuals."

According to Wilson, agencies with a single, distinctive culture are distinguished by highly scientific or technical work. Expertise is the dominant attribute of their culture. This furthers the argument that congress and the president's ex ante controls and similar measures to maintain accountability might not be as effective, at least with those agencies with a defined organizational culture, because they are far more resistant to political pressures. This hypothesis was tested and confirmed in part by Krause and Douglas' (2005) study that showed an agency's concern with its reputation supersedes its structure as an explanation for performance. Culture, however, is difficult to quantify and researchers struggle to find objective measures. Because of this difficulty, scholars have relied upon perceptions of agency "performance culture." These perceptions are most commonly taken from surveys of federal agency staff. A key question, then, is whether these subjective perceptions of agency culture and performance are associated with objective measures of the same concepts. The answer, according to Walker and Boyne (2004), is yes. In their analysis of multiple, related studies, they found strong, statistically significant correlations between subjective and objective measures of the many dimensions of culture and performance (see also Brewer 2005; Wall et al. 2004; Bommer et al. 1995).

> *Hypothesis 6. If agency staff have positive perceptions about their agency's performance culture, the agency will have a higher PART score.*

As agency missions and activities vary so do their ideologies or policy preferences. Agency ideology is often measured by using estimations of their policy preferences, discussed in greater detail below. Agency ideology or preferences are of interest to scholars in exploring questions about how agencies preferences are aligned or differ from the president and congress and, therefore, shape expectations for how agencies perform (Bertelli et al. 2008; Clinton and Lewis 2007; Nixon 2004).   As discussed above, presidents appoint like-minded officials to head agencies in the hopes that they will help the president carry out his policy agenda. If this is the case, what happens to agencies whose preferences or ideologies are at odds with the president? The PART system was developed and utilized exclusively during the Bush administration, which means more liberal agencies may be disadvantaged when having their performance assessed.

Hypothesis 7: Agencies with more liberal ideologies will have lower performance scores.

While not as popular within the political science literature's examination of performance, internal agency characteristics as predictors of performance are strongly rooted in public policy and public administration theory. While I am not arguing that internal characteristics necessarily offer a superior explanation for agency performance, they certainly deserve to be a part of a comprehensive model of performance alongside external characteristics.

In Chapter 3 I specify how I measure each of the variables and revisit my expectations for how each independent variable affects performance. I run the external and internal models separately and then report results for each as well as the combined model.

CHAPTER THREE

Quantitative Analysis

Chapter Preview:

In this chapter I explain the data, methodology and results of the quantitative analysis. I perform my empirical analysis in stages. First I examine the impact of external-only characteristics on performance followed by internal-only characteristics. I then run a model to determine how well the two sets of variables fare when combined. The questions are to what extent do internal and external variables explain the variance in PART scores and which variables are the most significant predictors of performance?

Data:

I employ a two-part methodology to answer what impact external and internal variables have on agency performance. The large-N quantitative analysis of the independent variables in this chapter will capture both internal and external agency characteristics. The analysis will consist of three models. The first will include only the external characteristics to test the effect on performance. The second model will include only internal characteristics and the third will incorporate both internal and external characteristics to determine the relative importance of each set of characteristics on performance. In Chapter 4 I offer a qualitative, exploratory analysis that includes a closer examination of agency performance based on additional factors like agency type, mission and culture.

Dependent Variable: PART Scores

The dependent variable used in this analysis is the Program Assessment Rating Tool (PART) score given to more than 1,000 federal agency programs between 2002 and 2008. As mentioned in the previous chapter, PART is a 25-part questionnaire completed by designated agency program staff in coordination with the OMB. The PART survey is made up of four sections, and each is weighted according to its importance in driving results.

After the PART score and rating are determined, program staff work with OMB to develop an improvement plan that outlines management actions the agency will take to improve performance. Most improvement plans will include action items to be completed within a year. Any agency progress toward meeting their improvement plan was updated twice yearly on ExpectMore.gov.

The majority of the programs rated between 2002 and 2008 scored ratings of adequate or better (see Table 3.1). Seventeen percent (173) of programs received a "results not demonstrated" (RND) rating and only two percent (26) of programs received an inadequate rating. Of the remaining programs, 29 percent (298) received an "adequate" rating, 32 percent (326) a "moderately effective" rating, and 19 percent (193) received an "effective" rating.

[Insert TABLE 3.1]

Scores from the results section matter most; they count for 50 percent of the total score. For example, while the Child Care Access Means Parents in School (DOE) and the Indian Health Service Resource and Patient Management System (HHS) programs received similar scores in the first three sections (80, 75, 90 and 80, 77, 100 respectively),

their overall ratings differ because of the lower score by the Child Care Access program (25 compared to 89) on the results section. The difference in scores on that one key section meant the Indian Health Service received the highest rating, "effective", while the Child Care Access program received only an "adequate."

I code the dependent variable three different ways. First, I use a categorical dependent variable (PART1) to capture which of the five possible ratings the program receives (results not demonstrated, inadequate, adequate, moderately effective or effective). The breakdown of the number of programs within each of those categories is in Table 3.1. Second, I use a dichotomous dependent variable (PART2) to indicate if a program is "performing" or "not performing." I use the Bush administration's classification for "performing" programs rated in the three highest categories (adequate, moderately effective and effective). Of the 1,016 programs rated using PART, 199 are not performing, while 816 programs are performing. Third I use a continuous dependent variable (PART3) on a 0-100 percent scale. This number is derived by multiplying the programs' scores in each of the four rating categories by the appropriate weight and then summing these scores for their total. PART3 scores range from 18.6 to 100 percent.

Independent Variables:

The independent variables are grouped according to the two different explanations I believe influence agency performance. Variables used to measure external characteristics affecting performance include "PERCENT APPOINTED," "UNIFIED" and "BUDGET." Each of these is a means through which the president and/or congress can control and influence agencies and their performance. PERCENT APPOINTED and UNIFIED are measured at the agency level, while BUDGET is measured at the program

level. Variables used to measure internal agency characteristics affecting performance

include agency IDEOLOGY, CULTURE, constituency SUPPORT and LEADERSHIP.

Each of the internal variables is measured at the agency level. Control variables include

the YEAR ASSESSED and PROGRAM TYPE to rule out any systematic differences in

scores based on the year evaluated or type of program. The summary statistics and

frequency distributions for the independent variables are found in Tables 3.2-3.3.

[Insert TABLES 3.2 and 3.3]

External Characteristics

Percent Appointed: Using Lewis' (2008) findings and his assertions about how

politicization affects performance, the measure PERCENT APPOINTED captures the

ratio of appointees to careerists in 28 agencies. I use Lewis' agency-level percent

appointed scores and apply them to all programs within the same agency (e.g., all

Department of Agriculture programs receive the same score). The scores range from a

low of .02 for the Department of Veterans Affairs to a high of 3.47 for the Department of

Education. These scores are based on 2004 data. Lewis (2008) hypothesizes that the

Department of Education, HUD and the Department of Labor are the most politicized

(i.e., have the most appointees v. careerists) because their policy commitments are more

democratic, therefore then-President George W. Bush responded by placing more of his

own appointees in those agencies.

Budget Size: The budgeting process is a powerful tool that the president and congress

use to control agency behavior, or at least that is frequently asserted. While attempts to

control agencies may yield better responsiveness, performance can be negatively affected

(Meier and Smith 1994). The assumption, although not directly tested in this dissertation,

is that the "thickening," or the additional layers of agency management created by various attempts to control agencies, has negative consequences for performance (Light 1995). Based on previous studies of PART however, I expect that agencies with larger budgets will perform better on PART because of the additional resources and staff they have available to implement management improvements (Gilmour and Lewis 2006; Boyne 2003). Agency program BUDGET is measured by a continuous variable indicating an individual program's 2008 fiscal year budget (in millions of dollars)[3] from the OMB's PART score spreadsheet dated September 2008. Budget size could have an independent effect on performance and, therefore, must be included in any analysis of agency program performance that includes programs of such varying size. There is a very large range of program budgets. In FY 2008, for example, budgets ranged from $2 million for the Department of Education's Comprehensive School Reform program to more than $500 billion for the Social Security Administration's Old-Age and Survivors Insurance program.

Unified versus Divided Government: Divided government can be measured two different ways. One way is to use a categorical variable to indicate if government was unified, split or divided when the agency was created. But since only 3 agencies were created during split government, I will instead use a dichotomous variable (UNIFIED) to indicate whether government is unified (e.g., the same party controls the presidency and both houses of Congress) or not. This measure is compatible with my hypothesis that divided government will negatively affect agency performance. This is an agency-level

---

[3] Ideally the budget would be measured by the FY program budget in the year prior to PART assessment. Unfortunately it is very difficult to track down the FY budgets for hundreds of programs dating back as far as 2002. To determine if using the FY 2008 budget was problematic, I created a subset of data containing only those programs rated in 2007 and 2008 using only the FY 2007 and 2008 budget numbers. There was no statistical difference in the bivariate or multivariate analyses.

variable that I created by researching the date the agency was created and then determining which party or parties controlled the White House and congress that year. In my dataset there are 17 agencies created during divided government, and 31 agencies created during unified government. All programs within an agency are assigned the same score.

Internal Characteristics

Data to measure agency CULTURE and LEADERSHIP are taken from the biennial Federal Human Capital Survey (FHCS), which captures the climate and culture within an agency and the aptitude for strong performance. First administered by the Office of Personnel Management (OPM) in 2002, the FCHS is designed to assess the attitudes and impressions of federal agency employees in four areas of their work experience with the aim of improving agency management and service delivery. OPM surveys the federal workforce every two years and asks a battery of questions within the four areas. Responses to questions are weighted to develop an index for each area. More than 210,000 federal agency staff from 37 agencies participated in 2008. Scores are available from the 2002, 2004, 2006 and 2008 surveys. As mentioned above, using this type of survey is based on prior performance research that finds perceptions of agency leadership, culture and performance are good proxies for more objective measures of these concepts.

The leadership index indicates how highly, or not, employees of an agency regard their leaders. *The performance index* indicates how much workers in an agency believe it promotes improvements in processes, products and services. *The talent index* indicates

the degree to which staffers think an agency has the talent needed to achieve its goals.
The job satisfaction index indicates how satisfied employees are with their jobs.

The OPM's FHCS measures are agency-level variables. All programs within a given agency are coded with the same score. The OPM provides indices and rankings for 2006 and 2008, but more than half of the agency programs were rated via PART prior to 2006. Because I am asserting that the FHCS scores have an effect on the PART scores, I will instead use selected questions from each of the four areas of the survey that are asked in 2002, 2004, 2006 and 2008 to be sure that I use the scores from the year prior to the program's PART assessment. This approach is similar to Lewis (2008), who used several questions he considered most relevant to the agency work climate, management and leadership. There is a major difference between Lewis' model and this analysis, however. Lewis used the FHCS data as his dependent variable to demonstrate that the management of an agency by an appointee or career staff affects performance (as measured by the FHCS). This analysis, by contrast, asserts that the FHCS taps staff attitudes and perceptions about their agency that will have an effect on performance (measured by PART scores).

The questions are used in all four years the survey has been administered. To ensure proper causal order, I will use responses from the survey prior to or occurring in the same year the program was rated via PART.

I use responses to the following questions or statements:

1. *In my organization, leaders generate high levels of motivation and commitment in the workforce. (LEADER1)*
2. *Managers communicate the goals and priorities of the organization (LEADER2).*
3. *I know how my work relates to the agency's goals and priorities.* (CULTURE1)

The specific questions were selected because they measure different dimensions of agency leadership and culture and have been previously examined in empirical studies of performance (see Brewer 2005). Culture and leadership are primary explanations for how well agencies perform and yet their roles have been somewhat overlooked in other examinations of performance. Leadership is also a very important driver of performance within the literature. The aspects of leadership which matter most are experience and expertise and the ability to foster cooperation among managers and staff. Questions 1 and 2 listed above measure staff perceptions of leadership (and become variables LEADER1 and LEADER2) and question 3 measures the degree to which staff believes their agency has a performance-oriented culture (CULTURE1).

Constituency Support: Constituency SUPPORT is measured by a July 2009 Gallup Survey of 9 federal agencies including: the Centers for Disease Control and Prevention (CDC), the National Aeronautics and Space Administration (NASA), the Federal Bureau of Investigation (FBI), the Central Intelligence Agency (CIA), the Department of Homeland Security, the Environmental Protection Agency (EPA), the Internal Revenue Service (IRS), the Food and Drug Administration (FDA), and the Federal Reserve Board. The survey was conducted by telephone and includes responses by more than 1,000 adults to the following question:

> *How would you rate the job being done by – [agency/department]? Would you say it is doing an excellent, good, only fair or poor job?*

Agency rankings based on responses to the above question can be found in TABLE 3.4.

[*Insert TABLE 3.4*]

With the exception of the Federal Reserve Board, the other agencies were previously
ranked according to the same survey question administered in 2003. The drawback of this
variable is its small n size. Because only 9 agencies were rated, there are only 123
observations when those agency-level scores are applied to the programs within those
agencies. I will, therefore, run the models with this variable and without it to determine
its value.

Agency Ideology: Bertelli et al (2008) developed a novel agency-level measure of
IDEOLOGY. The authors developed the 2007 *Survey on the Future of Government*,
which was administered to more than 6,500 appointed and career federal executives. The
survey was designed to obtain the "stated preferences" of the federal executives on key
votes in Congress in order to estimate ideal points and compare them to those of
legislators. The key votes were selected based on *National Journal's* list of key votes in
economics, social and foreign policy.   After obtaining the individual preferences, the
authors aggregated and averaged the individual preferences to create an agency-wide
ideology score.

This way of measuring agency ideology or preferences differs from previous
efforts that rely upon one of two methods, using expert judgments (Clinton and Lewis
2008) or using the partisan identification of the president, appointees or others to
determine agency preferences. With the first approach, scholars combine expert survey
results with objective data about agencies to classify agencies as liberal, moderate, or
conservative. These subjective judgments are based upon the broad mission or type of
agency (e.g., distributive, regulatory) or by the politics at the time of agency creation
(e.g., divided vs. unified government) (Gilmour and Lewis 2006). This approach

assumes that missions or agency types can be broadly classified as liberal or conservative. Further, if agencies are created during unified government, the agency's ideology will be closer to that of the party in control. The drawbacks of this approach are that it is subjective and lacks precision.

The other approach, using observed behavior, develops a measure for agency ideology by using the partisan identification of the president or appointees as a proxy. In total, Bertelli et al (2008) measured the ideologies of 33 agencies. These are admittedly not perfect scores. Some of the drawbacks include the lack of representativeness of the sample, and aggregating individual executive preferences into one agency score. These shortcomings aside, these scores help quantify the very abstract concept of agency ideology.

This is an agency-level variable. All programs within an agency will be coded with the same ideology score. The ideology scores range from the most liberal (the National Labor Relations Board) to the most conservative (the Department of Agriculture's Rural Development Program). A complete list of the agencies and their scores can be found in Table 3.5.

[*Insert TABLE 3.5*]

It is my expectation that ideology will be associated with performance. Because the PART program was administered during the Bush administration, I expect that programs within more liberal agencies will not perform as well as those programs within the most conservative agencies.

Control Variables:

Year Assessed: The year assessed is a categorical variable indicating the last year in which the agency program was assessed via PART. It is possible that some programs were previously assessed, but usually the original score was either ineffective or results not demonstrated. These original scores are not available so I cannot tell which programs have been rated previously. I am using the most recent score provided via PARTWeb. This measure will ensure that there are no systematic differences in programs scored in one year versus any other. There is some evidence that the year in which the PART assessment takes place matters. In his 2009 paper "Analyzing Federal Management Reform: PARTing the PART," Howard A. Smith finds that the number of programs in the performing category increases over time, which is statistically significant. Greitens and Joaquin (2009) found that programs scored in the first few years under PART received lower scores, suggesting that agency programs better understood how to comply with PART over time.

Program Type: Program type is measured by a series of dummy variables to indicate if an agency program is one of seven types: direct federal, credit, research and development, block/formula grant, competitive grant, capital assets, regulatory and mixed. Mixed is applicable to only one program and will therefore be treated as missing data.

There are numerous explanations within the literature for why certain types of programs are advantaged (e.g., those that distribute money) while others are disadvantaged (e.g., research and development programs). Research and development programs for example often do not yield short-term results and may then have lower

performance scores on PART because of the overwhelming focus on results. This argument is based on Frisco and Stalebrink's (2008) analysis of the difference between program/agency performance on PART versus GPRA. According to their comparison, GPRA allows research programs more time to produce their results, while PART does not. If this is true, then research programs should receive lower scores than other program types. There is also recent evidence that regulatory and redistributive programs are disadvantaged under PART because they are typically favored by Democrats and disfavored by Republicans (Greitens and Joaquin 2009). The relationship between program type and PART scores will be explored using bivariate and multivariate analysis in Chapter 3 and PART scores and agency type will be explored in Chapter 4.

Potential Limitations

As with any research endeavor, there are a number of potential limitations in my analysis of performance. The most noteworthy are measurement validity and the level of analysis.

Measurement Validity

Perhaps there are few large-N studies that quantify internal agency variables' affect on performance because concepts like culture, ideology and constituency support are difficult to measure. These concepts, unlike those tapped for the external variables, are subjective. I have had many challenges selecting valid measures of these variables. The leadership and culture variables I use from the Federal Human Capital Survey (FHCS) are too closely correlated and, therefore, cannot be used in the same model. Constituency support, or public support, for federal agencies is also difficult to measure. Public opinion of federal agencies tends to be captured by broader assessments of trust in

government and the bureaucracy at large. The Gallup Survey I use is a great measure

because it taps support for specific agencies. The drawback, however, is that it only asks

about support for nine agencies. I believe that the small n size of this variable skewed the

results of my model. I therefore decided to omit it from my results. Future analyses

should strive for more precise measures of these difficult concepts. These concepts will

be explored further in my qualitative chapter.

A related point about the "internal" variables I have chosen is that the

characteristics I have chosen to measure are not necessarily independent. For example,

ideology does not develop independently of the broader agency context. Each

characteristic is undoubtedly influenced by the agency's design, history, structure and

staff so I must clearly state that all of these variables are related to some extent. I once

again point to the strong relationship between the culture and leadership variables I

selected that prohibited both from being included in my empirical models.[4]

Level of Analysis: Agency versus Program

In this dissertation I discuss both agency and program performance sometimes

interchangeably, which reveals one of the many measurement challenges I encounter in

the empirical analysis. The broader literature uses terms like "bureaucratic" and

"agency" performance, while PART scores offer a more disaggregated unit of analysis,

which is the program level. The advantage of PART scores is that there are many more

units of analysis, more than 1,000 total. The difficulty, however, is that many of the

independent variables with the exception of program type, budget and year assessed, are

measured at the agency level. A related criticism of PART is that the scores fail to

---

[4] I tested for multicollinearity by examining the Variance Inflation Factor or VIF scores for each of the independent variables. These scores are found in Table 3.6. The individual VIF scores are all below 10, which means multicollinearity is not a major concern.

account for how programs perform in the larger agency context (GAO 2004). As 2004

testimony from Paul Posner of GAO states:

> **Disaggregating programs for PART purposes could ignore the interdependence of programs recognized by GPRA by artificially isolating programs from the larger contexts in which they operate.**

Certainly it would be possible to develop program-level measures of other

variables like ideology, culture, and percent appointed, but this is a very time-consuming

task and developing measures for each would be a dissertation unto itself.

The reason that selecting the appropriate level of analysis is important is that it

shapes and limits the extent to which findings are generalizeable. I offer many

explanations for "agency performance," but I must be clear that my findings are only

applicable to federal agency *program* performance.

Methodology:

Bivariate Analysis

As a first test of my hypotheses, I ran a series of bivariate associations between

the independent variables and PART scores. It is important to first revisit my

expectations for the relationship between these variables and PART scores. Beginning

with the internal variables, I expect both leadership (LEADER1) and culture

(CULTURE1) to have a positive affect on performance scores. Based on research,

leaders can motivate their staff to work harder and a strong culture can encourage

progress toward agency goals. Both of these variables are continuous so I ran simple

bivariate correlations to determine in what direction and how strong the relationship is

between LEADER1, CULTURE1, and PART scores. These associations are found in

Table 3.6. LEADER1 and LEADER2 are positively correlated with PART scores. Only

LEADER1, however, is significant. CULTURE1 actually has a negative coefficient, which is the opposite of my expectation and indicates a null relationship with PART scores. I will explore this finding further with the multivariate analysis. The size of LEADER1's coefficient (0.15) indicates a moderate, positive relationship between leadership and PART scores.

[Insert TABLE 3.6]

Agency IDEOLOGY and constituency SUPPORT should also be positively correlated with PART scores. My expectation for IDEOLOGY is that conservative agencies (ideology is scored on a range from the most liberal (-1) to the most conservative (+1)) receive higher PART scores because their preferences are closer to those of the president. Constituency SUPPORT is an important resource for agencies to use to withstand pressure from congress and the president. A positive coefficient indicates strong support. Both variables are continuous and their correlation coefficients are also found in TABLE 3.1. Similar to the LEADER1 variable, the coefficients for both IDEOLOGY and SUPPORT are positive and significant.

I also treat RESEARCH AND DEV as an internal variable, although I also consider it a control variable. The expectation based on critiques of PART, (see Frisco and Stalebrink 2008), is that research and development programs will be systematically disadvantaged by PART because results take longer to produce and this is not taken into account by PART, which evaluates most programs only one time. I use a series of dummy variables to indicate each type of program. In Table 3.7 I have listed the number of programs by type. The number of programs per category range from 34 credit programs to 321 direct federal programs.

[Insert TABLE 3.7]

In Table 3.8, I compared ratings for the different program types. The table includes the number of programs per type receiving each of the five ratings. Contrary to the critics of PART (Gueorguieva, et. al. 2008), which claim research and development programs will be disadvantaged due to the time it takes to achieve results compared to other program types, research and development programs appear to receive the highest percentage of effective and moderately effective ratings. Eighty percent of the research and development programs rated received a rating of either effective or moderately effective. Regulatory, Capital Assets and Direct Federal programs received similar percentages (69, 67 and 63, respectively) of effective and moderately effective ratings. Competitive Grant and Block/Formula Grant programs received the highest percentages (30, 34) of inadequate and results not demonstrated (or "not performing") programs. To be certain these differences between the PART scores of the different program types are significant however, I will use a difference of means test. This is the appropriate test since my independent variable is dichotomous. I specifically use a two-sample t test to determine whether the difference in means between research and development program scores and the scores of the other types is significant. The mean score of the research and development programs is 0.77 with a standard deviation of 0.12, while the mean score of the other programs is 0.71 with a standard deviation of 0.15. The difference in means is 0.06 with a t-statistic of 3.92 ($p < 0.001$). Of interest is that the mean PART score of the research and development program is higher than those of other programs' scores and this difference is statistically significant, contrary to my expectations.

[Insert TABLE 3.8]

I now move on to the external independent variables, which include PERCENT APPOINTED, BUDGET, and UNIFIED. PERCENT APPOINTED is a continuous variable so I can run a simple correlation with PART scores. Based on Lewis' (2008) findings, I expect the coefficient to be negative to indicate the greater the ratio of appointees will result in lower PART scores. The correlation coefficient (found in Table 3.4) is negatively correlated with PART scores and is highly significant. The size of the effect is also important. The coefficient is -0.18, which means that were the ratio of appointees to increase one percentage point, the PART score would decrease 18 percent. This is a large effect, but it will be interesting to see if the size is mitigated by the other variables in the multivariate analysis.

BUDGET is also a continuous variable (measured in millions of dollars) and I expect that an increase in budget will result in an increase in PART score. This is based on studies (see Gilmour and Lewis 2006), which find programs with larger budgets score higher, likely due to the additional resources agencies with more money can use to comply with PART and meet their goals. The correlation coefficient (also in Table 3.6) is positive, but small and insignificant. It does not appear, at least at this point, that budget is a predictor of PART scores.

The UNIFIED variable is dichotomous, so I will use a two-sample t test like I did to determine whether the mean PART scores of research and other types of programs were significant. I expect that agencies created during unified government will receive higher PART scores because they are given more discretion to act independently. The literature demonstrates that independence is a critical characteristic of a high performing agency (Potoski 1999). The mean PART score of the agencies created during unified

government (N=30) is 0.76 with a standard deviation of 0.12. The mean of agencies

created during divided government (N=18) is 0.71 with a standard deviation of 0.12. The

difference in mean PART scores is 0.05. The t-statistic is 1.4, but it is not significant (P

= 0.08). I therefore cannot reject the null hypothesis and claim that the difference in

PART scores is due to whether the agency was created during unified versus divided

government.

 The YEAR ASSESSED variable is important to include because I want to control

for any systematic differences in PART scores from one year to the next. Greitens and

Joaquin (2009) suggest that PART scores actually increased over time and that this could

be attributed to the number of programs being evaluated for the second time or to

agencies learning better how to comply with PART. Unfortunately it is not evident in the

dataset which programs were previously assessed, but a systematic and significant

increase in PART scores over time could support their findings. In Table 3.9 I have

included the mean PART score for each year from 2002 to 2008. With the exception of

2002, the mean scores are relatively stable from 2003-2008. This seems to suggest that a

great deal of learning took place after the first PART scores were given by OMB. To

determine if the difference in means between 2002 and the other years is significant, I

will once again run a two-sample t-test. The mean score for 2002 (N=49) is 0.67 with a

standard deviation of 0.18. The mean score for years 2003-2008 (N=793) is 0.72 with a

standard deviation of 0.14. The difference in mean scores (0.05) is significant at the 0.01

level, which means the scores were indeed lower in 2002.

<center>*[Insert TABLE 3.9]*</center>

<center>68</center>

Several of the variables have supported my hypotheses, but the bivariate analysis only tells us about the relationship or association between a particular independent variable and PART scores. Next I examine the relative effects of the internal and external variables on PART as well as the overall explanatory power of a model combining both sets of variables.

Multivariate Analysis

Because I want to compare the relative strengths of the external and internal models of agency performance, I will run each set of variables in a separate model and then I will run a combined model of all variables. These models are below.

External-Only Model: *y(PART Score) = b + b1(PercentAppointed) + b2(Unified) + b3(Actual08Budget) + b4(YearAssessed)*

Internal-Only Model: y(PART Score) = b +_b1(AgencyIdeology) + b2(Leader1) + b3(ResearchandDev) + b4(YearAssessed)

Combined Model: y(PART Score) = b + b1(PercentAppointed) + b2(Unified) + b3(Actual08Budget) + b4(YearAssessed) + b5(AgencyIdeology) + b6(Leader1) + b7(ResearchandDev)

I have coded my dependent variable three different ways and will, therefore, use the appropriate type of regression for each. PART1 is a categorical variable so I will use ordered probit. PART2 is a dichotomous variable so I will use logit regression and PART3 is continuous so I will use ordinary least squares (OLS) regression. From now on I will refer to the regression results according to the type of dependent variable (e.g., categorical).[5]

PART1 is coded 1-5 to indicate which of the five qualitative ratings an agency program received. The highest four ratings (effective, moderately effective, adequate,

---

[5] PART scores are program-level data, but I recognize that programs operate within a larger agency context. To address this important issue, I cluster programs by agency in each regression.

inadequate) are determined by the raw score (e.g., 85-100 percent = effective), while

RND are not. As I stated earlier, RND is given to programs that fail to establish OMB-accepted performance measures regardless of the overall score. Because this outcome category is different, I will first run the models without the 173 RND programs, then I will run the models including those programs. Because there is some subjectivity on behalf of OMB as to which programs receive RND, I will also perform a Heckman selection model to determine which specific independent variables might be determining if agencies receive this rating.

Results:

External-Only Model

The results of the external model are found in the first column of Tables 3.10-3.12. I will first review each hypothesis and discuss how well the results support my predictions. Once I review the results I will discuss the size and substantive effects of each significant independent variable on PART scores. In addition to whether the coefficients for the independent variables are in the expected direction and reach statistical significance, I need to more closely examine the size of each variable's effect on PART scores. I use the Clarify software developed by Gary King, Michael Tomz and Jason Wittenberg (2003), to assign specific values to a given independent variable while holding the values of all other variables constant, which can tell me the precise expected effect on PART scores. I will do this for each of my independent variables that achieved significance in one or more of the regressions and I will use actual agency program scores as examples. I acknowledge that while these examples are used to illustrate the potential for each of the significant independent variables to change the outcome of a

program's PART score or rating, I do not claim this would occur in every case. These examples, however, do show that a number of different variables have the potential to shape agency program performance.

Hypothesis 1 says, *as program budget size increases, program performance will improve.* The direction of the relationship between BUDGET and PART scores is as predicted in all of the regressions, but the coefficient never reaches significance. There is, therefore, no support for the hypothesis that an increase in budget will increase PART scores.

According to Hypothesis 2, *if government is divided at the time of agency creation, then the overall performance score will be lower.* The results for the dichotomous and continuous dependent variables (see Tables 3.11 and 3.12) support this hypothesis. Because the variable is coded 1 for unified government and 0 for divided or split government, the positive coefficient indicates that unified government is positively associated with higher performance. This result is significant at the 0.05 level for the continuous dependent variable PART3 and at the 0.01 level for the dichotomous dependent variable. Using the Clarify software, I input actual agency program scores to determine the substantive effect of the UNIFIED variable on PART scores. The Environmental Protection Agency was created during divided government which, as I hypothesize, should mean its programs do not perform as well on PART. I use the EPA's Global Change Research program to demonstrate the size of the effect of the UNIFIED variable on PART scores. The Global Change Research program was assessed in 2006 and received a PART score of 49 percent and a corresponding rating of "inadequate." This also means the program was considered to be "not performing" by the Bush

71

administration. The program's budget was $16 million and it is classified as a research and development program. Using the program's actual values for the other independent variables, I can compare the mean PART score for when the UNIFIED variable equals zero and when it equals one. In other words, will the PART score change simply by changing whether the agency was created during unified rather than divided government. The mean PART score increases from 0.72 to 0.75. This three percent change is important because it would have raised the Global Change Research program's actual PART score from 49 to 52 percent, which would classify the program as "adequate" or "performing."

Hypothesis 3 predicts that, *if the ratio of appointees to careerists is greater, then the overall performance score will be lower.* The results for the PERCENT APPOINTED variable meet this expectation. The coefficient is negative and highly significant in all of the regressions. As Lewis (2008) found, the higher percentage of appointed staff within an agency reduces agency performance due to increased politicization. PERCENT APPOINTED refers to the ratio of appointed to career staff within an agency. As Lewis (2008) found, an increase in this ratio seems to have a negative effect on agency performance as measured by PART scores. Using Clarify, I have used a specific agency program (the Department of Education's Federal Pell Grants program) to demonstrate how its PART score would have changed had its parent agency had more or fewer appointees. I do this by using the program's actual values for each of the variables (e.g. BUDGET) and then modify the value of PERCENT APPOINTED to isolate its effect on PART scores. I selected the Department of Education because it has the greatest ratio of appointees to careerists (3.47), and I use the Federal Pell Grants

program to demonstrate the strength of the PERCENT APPOINTED variable on PART

scores. The Federal Pell Grants program was assessed by PART in 2003 and received a

numerical score of 67 percent and a corresponding rating of "adequate." The program's

budget at the time was close to $12 billion. The Department of Education was created

during divided government (so UNIFIED = 0), the agency IDEOLOGY value is 0.05

(which categorizes it as moderate) and the LEADER1 score for the agency is 32 percent.

Using these values, I ran Clarify to compare the PART scores when PERCENT

APPOINTED equals the actual value of 3.47 versus several other possible values. If the

agency's ratio of appointees to careerists were 2.47 instead of 3.47, the mean PART score

for this particular program would increase from 59 percent to 62 percent. This is

certainly a sizeable increase. Using a more dramatic example, I set the PERCENT

APPOINTED value to 1.0. This time the mean PART score increases from 59 percent to

66 percent. A seven percent increase would be enough to move the Federal Pell Grant

program from the adequate to moderately effective rating category.

Internal-Only Model

Turning to the internal model, I ran the same series of regressions as I did for the

external variables. The findings are in the second column of Tables 3.10-3.12.

Strong leadership is expected to have a positive impact on performance.

Hypothesis 5 says that, *if agency staff have positive perceptions about their agency's*

leadership, the agency will have a higher PART score. The results are as expected.

LEADER1 has a strong, positive coefficient that is highly significant in all regressions.[6]

---

[6] Hypothesis 6 stated that agencies with positive staff perceptions about performance culture should have higher performance scores. The leader2 and culture1 variables have been dropped from the analysis due to problems with multicollinearity. Unfortunately I do not have culture and leadership variables that are independent from one another. Culture will be examined further in the qualitative analysis.

The LEADER1 variable has a very positive relationship with PART scores. My hypothesis is that strong, positive perceptions of leadership by an agency's staff will increase performance. The regression results support my hypothesis, but I want to estimate the size of this effect as I have for the other variables. The Department of Interior's Fish and Wildlife Service Habitat Conservation program was assessed in 2006 and received a score of 67 percent and a rating of "adequate." The LEADER1 value for the Department of Interior in 2006 was 0.3, which is rather low. Could an increase in this score have increased the program's PART score? The Department of the Interior was created during divided government and has an agency IDEOLOGY score of -0.2, and the ratio of appointed to career staff is 0.11. The program's annual budget is $95 million. Using these values, the estimated mean PART score is 0.67, but increasing the LEADER1 value to 0.4 or to 0.5 would increase the mean to 0.72 or 0.75, respectively. This is an increase of 5 percent in PART score for every 10 percent increase in LEADER1 value. Had the Fish and Wildlife Service Habitat Conservation program had a LEADER1 score of 0.4, the PART score would have been 72 percent and the program's rating would have been "moderately effective."

The agency IDEOLOGY variable is coded from Liberal (-1) to Conservative (+1). My expectation, according to Hypothesis 6, is that *agencies with more liberal ideologies will have lower performance scores.* Therefore, the coefficient should be positive to indicate that more conservative agencies score better on PART. The result supports my hypothesis. The coefficient is positive in all three regressions and highly significant in the OLS regression. Using Clarify I can assess their effects on PART scores using agency programs as examples. For agency IDEOLOGY I want to determine whether

changing the value of this variable from liberal to conservative or vice versa would

change the resulting PART score. I use the Department of Health and Human Services'

(HHS) Radiation and Exposure Screening and Education Program

 as my example. HHS is considered a liberal agency with an agency IDEOLOGY value

of -0.42. The agency was created during unified government (UNIFIED =1), received a

LEADER1 score of 0.39 and the ratio of appointed to career staff (PERCENT

APPOINTED) is 0.21. The Radiation and Exposure Screening and Education Program

has a budget of $2million. The program was assessed in 2006 and received a PART

score of 45 percent and a rating of "inadequate."

Using these particular values I can assess the estimated mean for the dependent

variable (PART scores) and compare it to the estimated mean when I change the value of

agency IDEOLOGY from -0.42 to 0.42. My hypothesis is that conservative agencies

perform better, so will changing the agency IDEOLOGY value to something more

conservative yield a higher PART score? The answer is yes. The estimated mean

increases from 0.74 to 0.80. A one point change in ideology yields a six percent increase

in PART scores. This six percent increase in PART score would have given the HHS

program a PART score of 51 percent and a rating of "adequate" and classified it as a

"performing" program.

The RESEARCH AND DEV is an interesting variable. Previous research

comparing GPRA and PART claims that research and development programs will not

score as well on PART because they do not yield short-term results (Gueorguieva, et. al.

2008). My regression results suggest the opposite is true because the coefficient is

positive and significant in all three regressions. I turn to Clarify to uncover whether a

program is classified as research and development can have a substantive effect on PART scores. I use the Department of Transportation's FAA Air Traffic Organization, Technical Operations program. The FAA Technical Operations program was assessed in 2008 and has an annual budget of $2.7 billion. The program received a PART score of 58 percent and a rating of "adequate." The Department of Transportation was created during unified government, has an ideology value of 0.08 (moderate to conservative leaning), a PERCENT APPOINTED value of 0.16, and LEADER1 value of 0.25 (or 25 percent positive). Using the actual values and then only changing the program type from direct federal to research and development, I find the mean of the PART score increases from 0.71 to 0.76. This is a sizeable increase and suggests that if the FAA Technical Operations program was a research and development program, the PART score would have been five percent higher (63 percent instead of 58 percent). This provides further evidence that research and development programs are *not* disadvantaged under PART. Rather, the opposite is true.

Next I turn to the combined model to determine whether these individual variable results hold and to find out if the overall explanatory power of the model increases by adding the variables together. I will also use the Clarify software to more precisely measure the substantive effects of the independent variables on PART scores.

Combined Model

The far-right columns of Tables 3.10-3.12 show the results of the combined model using the three different dependent variables and regressions. Many of the results are as expected. The direction and significance of several independent variables hold in the combined model. Looking first at the external variables, PERCENT APPOINTED is

negative and highly significant in all three regressions. The coefficient for the UNIFIED variable is in the expected direction in all three regressions, but only reaches significance with the dichotomous dependent variable. There is no support for the BUDGET variable in any of the regressions.

The coefficient for agency IDEOLOGY is positive and significant using the dichotomous and categorical dependent variables. The LEADER1 and RESEARCH AND DEV variables are positive and significant in all three regressions. The coefficient for YEAR ASSESSED, which serves as a control variable, is positive in all three regressions, but only reaches significance when using the categorical dependent variable.

Results Including "Results not Demonstrated" Programs

As I previously mentioned, I excluded RND programs from the first set of models because the rating is awarded differently than the other four categories. I want to determine if there are specific factors driving the OMB's decision to rate a program RND other than the given reasons that they lack acceptable performance measures and/or have failed to collect adequate date associated with those measures. First I run the same series of regressions now including those programs. Second, I will perform a Heckman selection model to determine if politics is playing a part in the assignment of that rating. Specifically I will use agency ideology and the percent appointed values to determine how each might affect a program receiving a rating of RND because these two variables measure aspects of the external and internal agency politics. I will also use year assessed since I previously found that program scores got better after 2002.

The results for the three regressions (categorical, dichotomous, and continuous dependent variables) are found in Tables 3.13-3.15. The strength and significance of

several variables remain similar. PERCENT APPOINTED still negatively impacts PART scores, which is highly significant using all three dependent variables. LEADER1 is also still significant and positively impacts PART scores. YEAR ASSESSED and RESEARCH programs are also positively associated with PART scores and their coefficients are significant in most regressions. Agency IDEOLOGY performs poorly and only reaches significance using the continuous dependent variable. UNIFIED does not reach significance in any regression.

[Insert TABLES 3.13-3.15]

Adding the RND programs into the analysis does have an effect as agency IDEOLOGY and UNIFIED no longer affect performance. The PERCENT APPOINTED, or the degree to which an agency is politicized still negatively affects performance, while LEADER1 still positively affects it. The year and program type also matter. I now perform a Heckman selection model, which isolates the RND programs so that I can better examine which specific variables are determining whether a program receives the RND rating. I believe that politics plays a part in this determination because OMB is afforded leeway to make these decisions. Therefore, I hypothesize that PERCENT APPOINTED and AGENCY IDEOLOGY will determine whether programs receive this rating. The larger the PERCENT APPOINTED, the more likely a program will receive an RND. The more liberal the ideology, the more likely the program will receive an RND since these ratings occurred during the Bush administration. I also include YEAR ASSESSED because a majority of the programs receiving RND did so in the first year PART was in effect.

The results of the Heckman model show that PERCENT APPOINTED is the primary variable determining whether programs receive an RND rating. The coefficient is negative and significant. The sign for the AGENCY IDEOLOGY coefficient is also negative, which indicates liberal programs are more likely to receive RND, but it is not significant.

Discussion

The results of the regressions support several of my hypotheses. Referring first to the analyses without RND programs, the external variables, PERCENT APPOINTED and UNIFIED, are important predictors of PART scores, while BUDGET appears not to play a role when accounting for other variables. All of the internal variables contribute to explaining agency performance at least in one or more of the regressions. Beyond the sign and significance of the coefficients for these variables I have examined their size and strength and found that each of these significant variables has the potential to change an agency's PART score. The empirical results, however, also show that neither external nor internal variables tell the whole story of agency performance. Instead it is the combination of variables, or external and internal agency characteristics, that give us a more complete picture of how and why some agencies perform better than others.

Beyond what has been learned from the models in this chapter, there is more variance to be explained. I was unable to empirically assess the role of agency culture, which I believe to be an important predictor of success. And because PART scores evaluated agency programs, rather than the overall agency, I did not provide a measure of agency type. There are many classic agency typologies. I review three and then discuss recent work applying one such typology to PART performance. I offer my own analysis

of agency type and PART and then move on to a discussion and analysis of agency

culture and performance.

TABLE 3.1 Number of PART Scores per Rating Category		
Rating Category	**Number of Programs within Category**	**Percentage of Programs within Category**
Effective	193	19
Moderately Effective	326	32
Adequate	298	26
Inadequate	26	2
Results not Demonstrated	117	17

Source: ExpectMore.gov

TABLE 3.2	**Summary Statistics for all Dependent and Independent Variables**				
Variable	**Obs.**	**Mean**	**Std. Deviation**	**Min**	**Max**
DEPENDENT VARIABLES					
PART1	1,015	3.3	1.3	1	5
PART2	1,015	.80	.40	0	1
PART3	1,015	.67	.18	.1	1.0
Internal Variables					
Leader1	955	.35	.06	.21	.55
Leader2	783	.58	.06	.33	.74
Culture1	955	.83	.03	.72	.92
Agency Ideology	908	-.05	.28	-.61	.49
External Variables					
Unified vs. Other	1,013	.62	.49	0	1
Percent Appointed	931	.68	.98	.02	3.5
Actual 2008 Budget	1,011	2644.26	23414.17	-177	505,062
Control Variables					
Year Assessed	1,015	4.0	1.6	1	7
Research and dev	1,009	.11	.32	0	1

TABLE 3.3 Frequency Distributions for Categorical Variables

Variable Name	Frequency	Percent	Cumulative
Year Assessed			
2002	62	6.10	6.10
2003	130	12.80	18.90
2004	203	19.98	38.88
2005	218	21.46	60.33
2006	220	21.65	81.99
2007	116	11.42	93.41
2008	67	6.59	100.00
Total:	**1,016**	**100.00**	
Program Type			
Block/Formula Grant	161	15.85	15.85
Capital Assets	85	8.37	24.21
Competitive Grant	177	17.42	41.63
Credit	39	3.84	45.47
Direct Federal	366	36.02	81.50
Mixed	1	0.10	81.59
Research and Dev.	113	11.12	92.72
Regulatory	74	7.28	100.00
Total:	**1,016**	**100.00**	
Rating			
RND	173	17.03	17.03
Ineffective	26	2.56	19.59
Adequate	298	29.33	48.92
Moderately Effective	326	32.09	81.00
Effective	193	19.00	100.00
Total:	**1,016**	**100.00**	

TABLE 3.4	Federal Agency Rankings of Constituency Support			
Agency Name	Excellent/Good %	Only Fair %	Poor %	No Opinion %
CDC	61	24	10	5
NASA	58	26	7	10
FBI	58	27	7	8
CIA	47	29	13	12
Department of Homeland Security	46	31	18	4
EPA	42	33	20	4
IRS	40	36	21	5
FDA	38	37	22	3
Federal Reserve Board	30	35	22	13

Source: Gallup Poll, July 2009.

TABLE 3.5	Rankings of Agency Ideology from Liberal to Conservative
Agency Name	**Ideology (Ranked from Liberal to Conservative)**
National Labor Relations Board	-0.68
Department of Health and Human Services – NIH	-0.61
National Science Foundation	-0.58
Environmental Protection Agency	-0.5
Department of Health and Human Services	-0.42
Federal Trade Commission	-0.41
National Archives and Records Administration	-0.35
Department of the Treasury	-0.27
Department of State	-0.24
Department of the Interior	-0.2
Department of Labor	-0.16
Department of Commerce	-0.13
Department of Commerce – Census	-0.13
Social Security Administration	-0.13
Department of Commerce – NOAA	-0.12
Department of Housing and Urban Development	-0.02
Department of Education	0.05
Department of Transportation	0.08
Department of Energy	0.1
Department of Veteran Affairs – Health Admin	0.15
Department of Agriculture	0.16
Department of Veteran Affairs	0.16
Nuclear Regulatory Commission	0.18
Department of Justice	0.19
General Services Administration	0.2
National Aeronautics and Space Administration	0.22
Department of Homeland Security	0.32
Department of Defense	0.37
Department of Agriculture – Agriculture Res. Service	0.4
Department of the Navy	0.43
Department of the Air Force	0.43
Department of the Army	0.45
Department of Agriculture – Rural Development	0.49

Source: Bertelli et al. 2008.

TABLE 3.6	VIF Scores for Explanatory Variables as Test of Multicollinearity	
Variable Name	**VIF**	**1/VIF**
Direct Federal	6.58	0.15
Competitive Grant	4.47	0.22
Block Formula	3.96	0.25
Research and Development	3.36	0.30
Five	3.30	0.30
Six	3.26	0.31
Four	3.15	0.32
Capital Assets	3.04	0.33
Seven	2.49	0.40
Leader1	2.47	0.41
Culture1	2.34	0.43
Regulatory	2.30	0.43
Leader2	2.00	0.50
Agency Ideology	1.33	0.75
Percent Appointed	1.22	0.82
Actual08 Budget	1.03	0.97
Mean VIF	**2.89**	

TABLE 3.7 Correlation Between Continuous Independent Variables and PART Scores	
Independent Variables	**Correlation Coefficient**
Agency Ideology	0.09**
Leader1	0.15***
Leader2	0.04
Culture1	-0.03
Const Support	0.25**
Percent Appointed	-0.18***
Actual08Budget	0.02
*** = .001, ** = .01	

TABLE 3.8	Number of Programs by Type
Program Type	
Block/Formula Grant	161
Capital Assets	85
Competitive Grant	176
Credit	40
Direct Federal	366
Research and Development	113
Regulatory	73
Mixed	1
Total:	1,016

TABLE 3.9	Number of PART Ratings by Program Type				
Program Type	Effective	Mod. Effective	Adequate	Inadequate	Results not Demonstrated
Block/Formula Grant	18	38	52	7	47
Capital Assets	18	32	22	3	10
Competitive Grant	18	46	59	9	44
Credit	4	7	12	1	5
Direct Federal	83	119	115	4	45
Research and Development	33	51	19	2	8
Regulatory	19	22	19	0	14
Chi2 = 55.11*** ***=0.001					

TABLE 3.10	Mean PART Score by Year Assessed
Year Assessed	Mean PART Score
2002	0.67
2003	0.72
2004	0.71
2005	0.71
2006	0.72
2007	0.73
2008	0.72

TABLE 3.11 Impact of Variables on PART Scores Minus RND (Ordered Probit)			
Independent Variables	External-Only	Internal-Only	Combined
Percent Appointed	-.17 (.05)***	-	-.16 (.05)***
Actual08Budget	1.5 (1.4)	-	2.7 (1.5)
Unifiedvsother	.16 (.08)	-	.13 (.09)
AgencyIdeology	-	.28 (.15)	.40 (.15)**
Leader1	-	4.0 (.73)***	2.9 (.78)***
ResearchandDev	-	.36 (.12)**	.25 (.12)*
YearAssessed	.06 (.02)*	.05 (.03)***	.05 (.03)*
Cut 1	-1.7 (.15)	-.28 (.29)	-.7 (.30)
Cut 2	-.03 (.13)	1.4 (.28)	.99 (.29)
Cut 3	1.0 (.13)	2.4 (.29)	2.1 (.30)
Pseudo R2	.01**	.03*	.03*
Chi Square	24.37***	45.91***	49.67***
N	753	729	711
*** = .001, ** = .01, * = .05			

TABLE 3.12 Impact of Variables on PART Scores Minus RND (Logit)			
Independent Variables	External-Only	Internal-Only	Combined
Percent Appointed	-.37 (.10)***	-	-.36 (.10)***
Actual08Budget	2.3 (3.7)	-	3.4 (4.3)
UnifiedversusOther	.40 (.16)**	-	.40 (.18)*
AgencyIdeology	-	.59 (.29)*	.76 (.30)**
Leader1	-	5.4 (1.4)***	3.3 (1.5)*
ResearchandDev	-	.80 (.25)***	.69 (.26)**
YearAssessed	.05 (.05)	. 04 (.05)***	.05 (.05)
Pseudo R2	.03*	.03*	.05*
Log Likelihood	-489.6	-471.81	-452.5
N	753	729	711
*** = .001, ** = .01, * = .05			

TABLE 3.13 Impact of Variables on PART Scores Minus RND (OLS Regression)			
Independent Variables	**External-Only**	**Internal-Only**	**Combined**
Percent Appointed	-.03 (.01)***	-	-.03 (.01)***
Actual08Budget	7.8 (1.9)	-	1.2 (8.4)
UnifiedVersusOther	.03(.01)*	-	.02 (.02)
AgencyIdeology	-	.05 (.04)	.07 (.04)
Leader1	-	.53 (.17)**	.39 (.18)*
ResearchandDev	-	.06 (.03)*	.05 (.02)*
YearAssessed	.01 (.00)**	.01 (.00)	.01 (.01)
F	9.98***	4.79**	40.42***
R2	.05*	.08	.11
N	753	729	711
*** = .001, ** = .01, * = .05			

TABLE 3.14 Impact of Variables on PART Scores (Ordered Probit)			
Independent Variables	**External-Only**	**Internal-Only**	**Combined**
Percent Appointed	-.23 (.07)***	-	-.21 (.05)***
Actual08Budget	8.8 (6.7)	-	8.9 (6.6)
Unifiedvsother	.20 (.18)	-	.17 (.13)
AgencyIdeology	-	.19 (.20)	.33 (.21)
Leader1	-	5.5 (1.6)***	4.0 (1.7)*
ResearchandDev	-	.55 (.18)**	.45 (.17)**
YearAssessed	.06 (.03)	.05 (.03)	.05 (.03)
Cut 1	-.80 (.18)	1.2 (.68)	.59 (.56)
Cut 2	-.70 (.19)	1.3 (.68)	.70 (.56)
Cut 3	.19 (.20)	2.2 (.70)	.16 (.58)
Cut 4	1.1 (.21)	3.2 (.67)	2.6 (.55)
Pseudo R2	.02	.04	.04
Chi Square	27.07***	34.25***	173.15***
N	927	897	876
*** = .001, ** = .01, * = .05			

TABLE 3.15 Impact of Variables on PART Scores (Logit)			
Independent Variables	**External-Only**	**Internal-Only**	**Combined**
Percent Appointed	-.44 (.09)***	-	-.41 (.06)***
Actual08Budget	.00 (.00)	-	.00 (.00)
UnifiedversusOther	.17 (.33)	-	-.06 (.31)
AgencyIdeology	-	.34 (.65)	.41 (.69)
Leader1	-	11.1 (3.8)**	8.7 (3.5)*
ResearchandDev	-	1.1 (.40)**	1.0 (.38)**
YearAssessed	.20 (.06)**	.22 (.08)**	.23 (.07)**
Pseudo R2	.07	.07	.10
Log Likelihood	-427.55	-408.24	-386.89
N	927	897	876
*** = .001, ** = .01, * = .05			

TA BLE 3.16 Impact of Variables on PART Scores (OLS Regression)			
Independent Variables	**External-Only**	**Internal-Only**	**Combined**
Percent Appointed	-.04 (.01)***	-	-.04 (.01)***
Actual08Budget	1.5 (1.3)	-	1.7 (9.3)*
UnifiedVersusOther	.03 (.03)	-	.03 (.02)
AgencyIdeology	-	.03 (.03)	.05 (.04)
Leader1	-	.89 (.25)***	.64 (.24)**
ResearchandDev	-	.10 (.03)**	.08 (.03)**
YearAssessed	.01(.01)*	.01 (.01)*	.01 (.01) *
F	10.69	10.82	40.57
R2	0.09***	0.12***	0.17***
N	927	897	876
*** = .001, ** = .01, * = .05			

CHAPTER FOUR

A Closer Look at Agency Performance: The Role of Agency Type and Culture

This chapter further explores explanations of agency performance not included in the empirical analysis. In chapter 3 I found several variables to have a significant effect on agency program performance. Confirming my expectations, an increase in the ratio of appointed to career staff had a negative effect on performance, while agencies with conservative ideologies, positive perceptions of leadership, and those created during unified government tended to have better performance scores on PART. Do these variables tell us everything about why some agencies perform better than others? While they offer new insights into performance, they do not explain all of the variance with respect to how well agencies fared on PART. For instance, the R-square for each OLS regression is modest and reaches only 0.24 for the combined model. This means much more variance in the dependent variable remains to be explained.

Chapter Preview:

In this chapter I explore agency type and culture as additional explanations of performance. Previous studies have pointed to agency type and culture as especially important predictors of performance (Ban 1995; Derthick 1990; Wilson 1989; Gormley 1986). Because PART scores are program-level variables, I accounted for the type of *program,* but not the type of agency. I considered including dummy variables for the type of agency, but because the small, independent agencies do not necessarily fit within one of the "standard" typologies discussed below, I felt nothing substantive would be

gained by incorporating agency type into those models. Agency type is important, however, which is why I discuss it in this chapter. Below I will discuss some classic agency typologies, highlight recent work that applied one of these typologies to PART performance, and offer my own analysis of agency performance based on agency type.

Following my analysis of agency type and performance, I move on to a discussion of the importance of agency culture. It was my hope to include a measure of agency culture in the empirical analysis because of its prominence in the literature and the lack of large-N empirical work testing culture's impact on performance. Unfortunately, as I note in the potential limitations section of Chapter 3, the measure I used was too highly correlated with the measure for leadership and was omitted from the models and I was unable to develop another suitable empirical measure. In this chapter I will define agency culture, discuss its importance in explaining agency performance and use a few selected illustrative case studies to demonstrate how agencies with different cultures performed on PART.

Since quantitative analysis has limitations, I engage in a case study to explore the impact of culture on agency performance. Case studies, according to King, Keohane and Verba (1994), can and should be focused, systematic comparisons of data across units. Case studies offer a way for me to explore my hypotheses for how agency type and culture affect performance through the "systematic collection of the same information across carefully selected units." I proceed first by identifying a few agencies whose culture vary. I determine their culture type by drawing upon a number of sources. Wilson's (1989) work from more than 20 years ago still seems to provide a reasonable approach to observing and measuring culture. I also draw upon Ban's (1995) analysis of

how culture type affects the roles of public managers. Each identifies some broad types of culture that shape my analysis. Second, I determine how to compare these cases on other data points that might impact their performance. These include variables from Chapter 3 as well as agency mission and agency type. Third, I examine how well the agencies fared on PART and offer an analysis of how culture may have played a role or not. This analysis is solely exploratory, so more empirical work is needed before generalizations can be made.

<u>Agency Type</u>

There are several classic agency typologies. I review three of the most prominent (Lowi 1964; Gormley 1986, Wilson 1989) and highlight what each offers in terms of how agency type might impact performance. I then discuss how Lowi's typology has recently been applied to the study of agency performance and PART, followed by my critique of this work and my own analysis of agency type and performance.

In his classic and often-cited 1964 review article, Theodore Lowi classified public policies according to one of four categories: distributive, redistributive, regulatory, and constituent. Based on these categories or types of policies, Lowi made predictions about the type of political relationships associated with each, which I believe can also help explain performance.

Distributive policies are those for which the government provides some type of good or service to a specific class of recipient. Distributive policies are often referred to as "pork barrel" policies and include farm subsidies and public land use. The benefits are concentrated, but the costs are diffuse. Distributive agencies include the Department of Agriculture and the Department of the Interior. Redistributive policies transfer resources

of some type from one group to another, usually through the collection of taxes. Examples of redistributive policies include social welfare programs administered by agencies such as the Department of Education and the Department of Health and Human Services. Regulatory policies are those that benefit many, but impose costs on a select few. Regulatory agencies like the Department of Energy develop policies and rules for specific groups, like corporations, to follow and impose sanctions for noncompliance. Examples include environmental and energy conservation policies. Lowi used constituent agencies as something of a catchall category for agencies that do not fit neatly within one of the other categories. There are also independent agencies such as the Department of Defense and the Office of Management and Budget.

Gormley (1986) analyzed regulatory policymaking, arguing that previous typologies including Lowi's did not account for two important characteristics: public salience and technical complexity. Public salience refers to how many people in the general public care about a given issue. If there are many people who care about an issue (broad scope of conflict) and with great intensity, public salience is high. If an issue is only of interest to a narrow group of people, or those affiliated with a particular industry, public salience is low. Technical complexity refers to an issue that raises questions requiring a high level of specialized knowledge to address (Gormley 1986, p. 598). Both complexity and salience can change over time, although complexity is less likely to do so. According to Gormley, the combination of public salience and technical complexity, whether each is high or low, will result in one of four different types of regulatory politics; board room politics (low salience, high complexity), street-level politics (low

salience, low complexity), hearing room politics (high salience, low complexity), and operating room politics (high salience, high complexity).

Public salience and technical complexity also affect who participates. "Regulars" include bureaucrats and officials from a particular industry. They always participate. Depending on salience and complexity, however, predictions can be made about which "irregulars" including politicians, citizens and interest groups may also participate. For example, politicians and citizens are usually attracted to salient issues, but not so much to complex ones. Salience and complexity also predict the "dominant participant." Politicians will dominate high salience, low complexity issues because of the potential electoral consequences. High-level bureaucrats will dominate high salience, high complexity issues because they are the experts. Low-level bureaucrats handle low salience, low complexity issues according to standard operating procedures, and business groups dominate low salience, high complexity issues due to economic motives (Gormley 1986, p. 615).

In chapter 1, I briefly described Wilson's (1989) typology of agencies. In contrast to Lowi, Wilson categorizes agencies based on how well their activities and results can be observed rather than by the nature of the policies they carry out. The four types of agencies are procedural, craft, coping, and production. In procedural agencies, managers can observe the activities of their staff, but not the eventual outcome. The best example of a procedural agency is the military during peacetime. The training and preparation is observable, but whether or not these activities are effective in achieving peace and security is unknown. Craft agencies are those with observable outcomes, but unobservable activities. This applies to enforcement agencies, like the Wage and Hour

division of the Department of Labor. The compliance officers are located all across the

country and are not directly observed by supervisors. The outcome of their efforts,

including legal complaints, could be observed and evaluated. Neither the activities nor

outcomes of coping agencies are observable. This type of agency or organization is

usually made up of street-level bureaucrats like police officers or even teachers.

Production agencies are those like the Social Security Administration that have

observable outputs and outcomes. Another example is the IRS, which can observe its

staff and the amount of money they collect (Wilson 1989, p. 160). Why, according to

Wilson, are these types of agencies important for our understanding of performance? The

type of agency determines the process by which managers hold staff accountable and

build and maintain morale. Managers in production agencies can observe both the

activities and outcomes produced so they are better able to determine staff contributions

to results achieved. In a procedural organization, managers closely watch activities and

processes because outcomes cannot be observed. The focus therefore is on process, not

results, which could hamper overall performance on PART that stresses outcomes over

outputs. Craft organizations have observable results, but the staff conducts their daily

activities without supervision. The potential to shirk is great, so to be successful,

agencies of this type must foster a strong sense of mission to motivate staff to do their job

according to outlined rules and procedures. Coping organizations or agencies are the

most difficult to manage because neither outcomes nor outputs are observable. Poor

outcomes sometimes signal managers to look into and correct a problem, but these

"alarms" are often random and it is difficult to identify the source of the problem (Wilson

1989, 159-175).

The type of agency may also determine how susceptible agencies are to external control by the president and congress. Agencies like production and craft agencies with observable outcomes are more likely to be externally controlled than procedural and coping agencies, whose outcomes are not observed. Wilson provides case study examples for each type of agency and makes general predictions about how the type of agency and the strength of its mission could affect their success. I will revisit his predictions in the next section of the chapter when I discuss agency culture.

PART scores are measured at the program level, but because I want to determine how additional factors like agency type affect performance, I have created an average PART score for each agency, which is the sum of all agency program scores divided by the number of programs within the agency rated via PART. The average agency rankings by PART score are found in Table 4.1.

[Insert TABLE 4.1]

As you will see from looking at Table 4.1, the average agency PART scores range from a low of 35 percent for the Federal Communications Commission to a high of 94 percent for the Smithsonian Institution.   It is difficult, however, to make any real inferences from such a large table so the next step is to divide agencies according to different criteria to determine whether factors like agency type may help explain PART scores.

I turn now to a very recent article that applies Lowi's typology to agency performance on PART. Greitens and Joaquin (2010) evaluated how well different types of federal agency programs fared on PART. They asked two research questions. First, would agency program type affect the PART scores in each of the four sections: Program

Purpose and Design, Strategic Planning, Management and Results. Second, would there be greater variability in PART scores for redistributive and regulatory programs because their outcomes are more difficult to consistently measure. They hypothesized that redistributive and regulatory agency programs would receive lower PART scores because these programs were likely to be at odds with President Bush's agenda. Agency programs were categorized into the four Lowi types: constituent, distributive, regulatory and redistributive. Using PART's categorization of seven different programs types, Greitens and Joaquin made determinations as to which PART program types belonged in each of Lowi's four overall agency types. The coding from PART program types into Lowi's types is found in Table 4.2.

[Insert TABLE 4.2]

Greitens and Joaquin used t-tests to examine the difference in mean PART scores for each of the four types of agency programs. They also employed point-biseriel correlations to determine the strength and direction of the relationships between PART scores and program type. They found a large and significant difference in PART scores between redistributive and non-redistributive programs and that redistributive programs consistently receive the lowest PART scores among all agency program types, although the correlation is modest at best. They also found a greater variability in PART scores for redistributive and regulatory programs and attribute this variability to a greater difficulty in measuring outcomes for these program types. This finding ties into criticisms that OMB graders were inconsistent in their scoring and that it could be because of political motivations.

According to Greitens and Joaquin, constituent and distributive agency program types received higher PART scores than redistributive and regulatory types. One possible explanation is that their missions and goals are the most straightforward. Clear, unambiguous mission statements and goals were considered key determinants of performance by Wilson (1989) and this explanation has since been empirically validated (Chun and Rainey 2005).

For this reason, I choose to compare the types of agencies based on their average PART score (see Table 4.1) to see try and confirm Greitens and Joaquin's (2010) findings. Table 4.3 shows the average PART scores for each agency and an average PART score for agencies of each type. Interestingly, constituent agencies have the highest average PART score (74 percent), while distributive and regulatory agencies have nearly the same average (67 and 69 percent respectively) and redistributive agencies score the lowest on PART (60 percent). To determine whether the difference in averages is statistically significant, I conduct a two-sample t-test comparing the mean PART scores for programs in constituent agencies versus those in all other agencies. The mean PART score for all constituent agency programs is 0.74 with a standard deviation of 0.17. The mean PART score for all programs in the other three agency types is 0.64 with a standard deviation of 0.18. The t-statistic is 7.38 and the difference in means is 0.10 (p > t = 0.0000), which means the mean PART score of constituent agency programs is higher than in other types of agencies and this difference is statistically significant. These findings are partial confirmation of Greitens and Joaquin's analysis and suggest that more work should be done to uncover how agency type affects performance.

[Insert Table 4.3]

Because constituent agencies are consistently the highest performers whether I use the percent effective rating or average agency PART score, I want to know what makes these agencies different from the other three types. I begin first by applying what I learned in Chapter 3 about which variables seem to influence PART scores. In Table 4.4 I have listed each of the constituent agencies as well as their values on variables found to have a substantive and significant effect on PART scores from Chapter 3. I have eliminated the variables YEAR ASSESSED and PROGRAM TYPE because they are program level variables and not applicable here. Included are ideology, leadership,[7] and percent appointed scores for the agencies. Most of these agencies were created during unified government, but an average score is not applicable because this is a dichotomous variable. What can be learned from this table? Below I compare these four broad types of agencies across key explanatory variables from chapter three.

[Insert Table 4.4]

Beginning with constituent agencies, it appears that leadership is not driving the constituent agencies' high overall PART scores as the average leadership rating is below 50 percent. The average ideology score leans conservative, which as I found in Chapter 3 is associated with higher performance. This is one possible explanation for the category's higher ratings. The percent appointed average is also relatively low, which might also explain the strong performance of these agencies as a higher ratio of appointees was found to have a negative effect on performance. Next, I compare how the other three types of agencies fared using their average values for the variables found to be significant in Chapter 3.

---

[7] For each agency I have averaged the LEADER1 scores from each of the four years that the agency participated and received a score in the Federal Human Capital Survey from which the variable is drawn.

In Tables 4.5-4.7, I list the values of the significant independent variables from Chapter 3 for regulatory, distributive, and redistributive agencies. Interestingly, there are no major differences among distributive and regulatory agencies when their average scores are compared to those of constituent agencies. The average leadership scores are very similar and only distributive agencies are slightly less conservative (0.01) than regulatory (0.13) and constituent agencies (0.07). The average percent appointed values are also similar. Redistributive agencies, which have the lowest average PART scores, are also the most liberal (-0.11) and have the greatest percent of appointees (0.1). I would have to say that it is not the constituent agencies that stand out from this descriptive analysis, but the redistributive agencies, for how differently they score on those two variables.

[Insert TABLES 4.5-4.7]

We have learned that the type of agency potentially affects performance on PART, which builds on our understanding from the models in Chapter 3 that found agency politicization, leadership, politics at the time of agency creation, ideology, and program type matter. There is still more to be explained, however, about why agencies perform differently. Next I look at agency culture. This is an important concept and has proven quite challenging to measure, which is why it is frequently omitted in large-N analyses of performance.

Agency Culture

An agency's culture is also a commonly cited explanation for agency performance (Lewis 2003; Smith and Sosin 2001; Brewer and Selden 2000; Moon 2000; Mahler 1997; Ban 1995; Wilson 1989). According to Wilson (1989), culture is defined as the

"persistent, patterned way of thinking about the tasks of the agency, how staff interact, and how appropriate actions, rewards and consequences are defined."   This is echoed by Ban (1995), who finds that agency size, history, and mission shape culture. For Wilson, mission is very important to the development of culture as are defining the agency's "critical task," and an agency's degree of autonomy.

In addition to agency size, history and mission, leadership is an especially important component of agency culture. According to the Partnership for Public Service and American University's Institute for the Study of Public Policy Implementation (ISPPI), culture can be measured by "the extent to which employees believe that management ensures they have the necessary skills and abilities to do their jobs, is successful at hiring new employees with the necessary skills to help the organization, and works to achieve the organizational goals with targeted personnel strategies and performance management" (Partnership for Public Service 2009). It is with these components of culture in mind that I explore the relationship between agency culture and performance.

The agency's critical task is different from the agency's goals, which, as Wilson points out, can be numerous and vague. Tasks, rather, are the means agencies use to achieve their goals. For example, the primary goal of the Social Security Administration is to "deliver social security services to meet the changing needs of the public" (SSA.gov 2010). The task for the agency then, is to pay benefits on time and accurately. The SSA's task of paying benefits on time is a means to achieving its goal. When the task(s) of an agency are widely agreed upon and endorsed by staff, Wilson says the agency has a "sense of mission" (Wilson 1989 26). The agency's sense of mission encourages pride

among staff and enthusiasm for their work and ultimately helps the agency perform its tasks more effectively. The third organizational issue, agency autonomy, means that the agency has some freedom to redefine its tasks as necessary and that there is a sufficient base of external political support to draw upon when needed.

The task environment, as highlighted in James D. Thompson's classic book *Organizations in Action* (1967), varies greatly from one agency to another. No two task environments are alike and, therefore, their influence on the agencies' culture will vary as well. This has profound implications for agency performance depending upon the complexity and size of the organization. The task environment refers to the particular job the agency is assigned to do as well as the clients it serves, resources it has and the broader environment in which it operates. The task environment, according to Thompson shapes the agencies' culture.

What is agency culture? When agency staff agree upon and embrace their mission, Wilson says it becomes the organization's culture. When there is no such agreement, agencies tend to have multiple and conflicting cultures. This is the case for most agencies. According to Ban (1995), no pure culture type exists among federal agencies. Instead it is the extent to which there is a consensus within the subcultures that matters for behavior and performance. Achieving consensus is possible only when the goals are clear, which is also rarely the case.

Goal ambiguity creates culture conflict (Chun and Rainey 2005). According to Chun and Rainey (2005), federal agencies tend to have a high level of goal ambiguity, which means staff do not clearly understand what the work they do is meant to achieve. Arguably, goal ambiguity has implications for how well agencies perform, although there

is mixed empirical evidence (Weiss and Piderit 1999; Baum, Locke and Kirkpatrick 1998; West 1997; Lowi 1979). In a two-part empirical analysis, Chun and Rainey developed four measures of goal ambiguity and tested their relationship with a number of variables including work satisfaction, organizational structure and performance, to determine whether government performance efforts like GPRA and PART make the correct assumption that clear goals and missions are linked to strong performance.

Chun and Rainey (2005 531) define goal ambiguity as the "extent to which an organizational goal or set of goals allows leeway for interpretation, when the organizational goal represents the desired future state of the organization." As I noted above with examples, most federal agencies have ambiguous goals although the degree of vagueness varies. The different measures of goal ambiguity include mission comprehension ambiguity, directive goal ambiguity, evaluative goal ambiguity and priority goal ambiguity. They found that three of the four measures (excluding mission comprehension ambiguity) performed well and that organizational performance, as measured by perceived effectiveness on the 2000 U.S. Federal Employee Survey, was negatively impacted as goal ambiguity increased.

Using Wilson's (1989) and Ban's (1995) broad classifications of agency culture, Chun and Rainey's findings, and agency rankings on a federal survey about performance culture as selection criteria, I now examine four agencies as case studies to examine how culture affects performance. I expect that agencies with strong missions, effective leadership, clear goals and those whose staff perceive a strong culture will perform better than those without them. I establish how each agency measures up on these characteristics. First, I rank agencies according to both the the 2008 Federal Human

Capital Survey (FHCS) performance-oriented culture index and the 2009 *Best Places to Work* survey administered by the Partnership for Public Service.[8]   The 2008 FHCS ranked participating federal agencies according to four indices: leadership and knowledge management, performance-oriented culture, talent management, and job satisfaction. In Chapter 2 I discussed how these indices were developed, but I will provide a quick review for reference. The FHCS is designed to assess the attitudes and impressions of federal agency employees in four areas of their work experience toward improving agency management and their provision of services. OPM surveys the federal workforce every two years and asks a battery of questions within the four areas.   Responses to questions are weighted to develop an index for each area. More than 210,000 federal agency staff from 37 agencies participated in 2008. I am most interested in the performance-oriented culture index because it measures how much workers in an agency believe it promotes improvements in processes, products and services. In other words, it asks agency staff whether they believe their agency's culture promotes strong performance. The rankings for 32 agencies according to the performance-oriented culture index are found in Table 4.8. The index scores range from a low of just under 41 percent for the Broadcasting Board of Governors to a high of nearly 66 percent for the National Science Foundation.

[Insert TABLE 4.8]

The *Best Places to Work* survey is also administered every two years and was developed by the Partnership for Public Service and American University's Institute for the Study of Public Policy Implementation (ISPPI). I use the 2007 scores for agencies'

---

[8] This was supposed to be my empirical measure of agency culture, but I was unable to use it in Chapter 3 because of the high degree of correlation between my leadership and culture variables, both of which were from FHCS.

"strategic management," which measures how well staff believe managers successfully

hire employees with the required skills to do their jobs well and those who will work to

achieve organizational goals. It also measures how well managers have "targeted

personnel strategies and performance management" (Partnership 2009). I chose to use

the 2007 scores because the survey was administered prior to the end of the Bush

administration. I use the *Best Places* survey to validate the rankings of the FHCS. If

rankings for a particular agency are similar, then I can assume these are valid measures of

culture. The rankings according to the *Best Places* survey are found in the far right

column of Table 4.8.

Despite a few exceptions like the Social Security Administration and the National

Credit Union Administration, the scores are quite similar and the rankings of agencies

change little from one survey to the other. Given the similar scores, what might explain

the variation in agency culture and what impact, if any, do these scores potentially have

for how well agencies score on PART? I attempt to answer these two questions below by

examining select agencies as case studies. These cases are exploratory and, therefore, I

cannot make generalizations from my findings, but I am hopeful they will provide

insights for how culture might affect performance.

The FHCS measures federal agency staff perceptions of their agency's

"performance-oriented culture."   As I mentioned earlier in the chapter, there are several

factors that influence how well the so-called "performance-orientation" may or may not

become a part of an agency's culture. Among them are the strategic management of top

leadership and the agency's ability to synthesize and use the performance information to

make decisions (Trahant 2008; Newcomer 2006). In a survey analysis of how well

performance-orientation has become part of the management culture of federal agencies, Newcomer (2006) found that more than half of all managers reported performance information was used to guide decision making. Newcomer's results also suggest that program managers are learning about how well their programs work from data and program performance evaluations and that they feel they are held accountable for results. Despite these advancements toward performance-oriented cultures, there are challenges which agencies must overcome in their efforts to develop such a culture.

Among these challenges is deciding how to concisely capture an agency's mission with clear and measurable objectives (Trahant 2008). Deciding upon measurable program outcomes is another difficult. Agency managers have to figure out how to translate the organization's goals into individual staff job descriptions (Trahant 2008). Given these challenges, it is not surprising that agencies vary with respect to how well they have developed a "performance-oriented culture." Next I examine the different cultures in four agencies.

I have selected four agencies to provide a variety of size, mission, rankings on the FHCS and *Best Places to* Work surveys and their classifications per Wilson (1989) and Ban (1995). I hypothesize that agencies with well-defined missions and clear goals will perform best. The agencies selected are the Social Security Administration, Department of the Navy, NASA and the Department of Education. SSA and Navy were selected because Wilson contrasted the two agencies in his discussion of organizational culture. The SSA was discussed as an example of an agency with a very strong sense of mission and culture, while the Department of the Navy was used as an example of an agency with many competing subcultures. Ban also used the Department of the Navy as an example

of an agency with a "great diversity of organizational missions and structures," and she

concluded that these different cultures have made effective management difficult (1995

26). NASA and the Department of Education were selected because of their rankings on

the FHCS. NASA scored very high, while the Department of Education was in the

bottom third. I will examine how each agency scored on PART and each of my

independent variables, then look closely at their mission and goal statements to figure out

why these agencies perform differently. It is my hypothesis that a higher ranking on the

FHCS will be associated with higher performance on PART. It is my second hypothesis

that the SSA, which is considered a "strong culture" agency, will perform better than the

other agencies which have mixed cultures.

Social Security Administration

The Social Security Administration (SSA) was initially created as the Social

Security Board (SSB) in 1935 when President Franklin Roosevelt signed the Social

Security Act into law. The SSB had a rocky start with few staff and a minimal budget.

In 1939, the SSB lost its independent agency status and was consumed by the new

Federal Security Agency along with the Public Health Service, the Office of Education,

the Civilian Conservation Corp., and the U.S. Employment Service. In 1946, the SSB

was renamed the Social Security Administration, and in 1953 it became part of the new

Health, Education and Welfare Agency (HEW). HEW became the Department of Health

and Human Services, which housed the SSA until it once again became an independent

agency in 1995 (SSA.gov 2010).

Today the SSA employs more than 62,000 staff in 1,400 offices around the

country (OMB 2009). As I mentioned above, one of the SSA's most critical tasks is to

pay retirement benefits on time and accurately. Because this is an easily defined and agreed upon task, the SSA's retirement program, according to Wilson, has a strong sense of mission (Wilson 1989). The strong mission shaped the development of the SSA's service culture, which I argue is an important explanation for why the agency performs well. It is worth noting, however, that the SSA has struggled to maintain its strong sense of mission. When the Supplemental Security Income and Disability Insurance programs were placed in the SSA, the different nature of the programs' tasks threatened the agency's strong mission. It had to work very hard to absorb these new programs and tasks while trying to maintain its mission and culture (Derthick 1990). I look next at how well the SSA performs today by PART criteria.

The SSA's average PART score is a very respectable 76 percent and this score would place it in the "moderately effective" category. Beyond the assumption that the strong sense of mission and culture contributed to the agency's strong performance, I also want to look at how well it scored on other key variables from my empirical analysis. The agency's ideology is -0.13, which makes it liberal-leaning. The leadership score is 43 percent and the percent appointed value is 0.03. The ideology and leadership values would predict lower PART scores according to my earlier findings, but the low percent appointed value suggests the opposite. In addition, the SSA was created during unified government, which also predicts better performance. Since the values on these key independent variables offer mixed predictions about performance, I am led to believe something else is contributing to the SSA's strong performance.

I ran Clarify as I did in Chapter 3 to generate a predicted PART score based on the agency's average values (sum of program values divided by the number of programs).

The predicted PART score is 0.80 (95% CI = 0.67 <= μ_1 - μ_2 <= 0.94), while the actual

agency PART score based on the average of all program scores is 0.76. It appears that

the variables included in the empirical analyses do a good job of explaining the SSA's

performance on PART. This does not mean that culture does not have a place among

these variables. It is likely that culture is what shapes many of these variables such as

leadership and ideology, which in turn influence agency performance.

Department of the Navy

Wilson discussed the Department of the Navy at length and described it as an

agency with competing cultures. For this reason, I have selected the agency as a case

study to determine whether these different and diverse cultures impede performance.

The Department of the Navy was established in 1775. In 1949, the Department of

Defense (DOD) was established and the Secretary of Defense was given control over all

branches of the military. DOD employs more than half of all civilian employees, making

it the largest agency in my analysis (Ban 1995). With the Navy's size come many

different cultures and these cultures have implications for how staff do their jobs and,

therefore, how the agency performs. The Navy's mission is "to maintain, train and equip

combat-ready Naval forces capable of winning wars, deterring aggression and

maintaining freedom of the seas" (Navy.mil). This is straightforward, but as Ban (1995)

points out, the mission has to be constantly reevaluated in the face of changing world

politics. In addition, the methods for achieving such a mission differ based on whether

staff are active duty servicemen or civilians. For example, within the Department of the

Navy, there are pilots and seaman, and as Wilson (1989) points out, their cultures are

different and sometimes collide. The important question is whether or not this cultural

divide impacts performance.

The Department of the Navy's average PART score is nearly 77 percent[9], which

is similar to the SSA's. The Navy, by virtue of its place within DOD, is a constituent

agency, which appears to be advantageous for performance. Looking at the key

independent variables from Chapter 3, there is mixed evidence of why the Navy (and

DOD) performs well on PART. Using DOD's scores for agency-level variables, DOD

was created during divided government and perceptions of effective leadership are

modest at 40 percent. The values of these variables should disadvantage the agency, but

the percent appointed score is very low (0.04) and it is not surprisingly a conservative

agency (0.43), which might explain its strong overall PART score. I ran Clarify to

generate a predicted agency PART score based on the average values of the key variables

from Chapter 3. The predicted PART score is 0.74 (95% CI $= 0.62 <= \mu_1 - \mu_2 <= 0.85$)

or 74 percent, which is lower that what the Navy actually received, but the average score

is within the confidence interval. DOD ranked in the middle of agencies on the FHCS

performance-oriented culture index, so it is unclear whether culture explains the strong

performance on PART. Whether culture played a role in the DOD or Navy's PART

scores is unclear, but some might argue that the Navy and DOD would be advantaged on

PART because of the agency's major contributions to the wars in the Middle East and the

strong backing of the agency by then-President George W. Bush. This would suggest

that PART scores were, as critics argued, politically motivated. It could also be that for

agencies with a well-defined culture, culture is a strong predictor of performance, while

---

[9] The Navy's PART score was calculated by the average of the nine Navy programs that received PART scores. For OMB, the Department of Defense is considered the "agency."

for agencies like DOD with varied and competing cultures, the affect on performance is mitigated by other factors. I will discuss the potential for culture to have a mediating role in performance in the final section of this chapter.

NASA

NASA's mission is "to pioneer the future in space exploration, scientific discovery and aeronautics research" (NASA.gov). NASA was created in 1958 by President Eisenhower, who was reacting to the former Soviet Union's launch of an artificial satellite. Previously known as the National Advisory Committee on Aeronautics (NACA), NASA's focus in the 1960s and 1970s was getting men and technology into space. Later NASA's focus became the development and evolution of the space shuttle and international space station. NASA technology is also used to improve airline safety and health, among many other things. The agency is housed in Washington, DC with 10 field offices around the country (NASA.gov).

According to a 2007 report by the National Academy of Public Administration (NAPA), NASA struggles to "accomplish new missions and expand space exploration efforts" (Ballenstedt 2007). The report was conducted at the request of the Senate Appropriations Committee, which was concerned about how management and workforce challenges were affecting agency performance. NASA employs nearly 60,000 civil servants and contractors, a large workforce developed in response to previous missions. The agency has struggled to adapt to new missions while keeping its current structure intact (Ballenstedt 2007).

Despite some of these mission-related challenges, NASA programs scored very well on PART and the agency ranks highly on the FHCS and Best Places to Work

surveys. NASA had 13 programs evaluated by PART and scored an average of 76

percent. NASA is both a constituent agency and a conservative agency (0.22), which

should yield better performance on PART. In addition, NASA has a low percent

appointed score (0.08), which should also be associated with strong performance. The

agency was created during divided government and the leadership score is modest, which

could both negatively affect performance. What is interesting, however, is that NASA

ranks fourth overall on the FHCS performance-oriented culture index. The perception of

a strong culture might explain the agency's relative success on PART.

The predicted PART score generated via Clarify is 0.69 (95% CI = 0.60 <= μ_1 -

μ_2 <= 0.79), while the actual average agency PART score is 0.76. The average PART

score is within the confidence interval, while there is a difference in the raw scores, the

model seems to explain NASA's performance relatively well.

Department of Education

Unlike the other three agency cases, the Department of Education did not fare as

well on PART. The agency's average score for the 93 programs evaluated is only 50

percent. This score is significantly lower than the SSA, DOD and NASA. The

Department of Education was established by the Department of Education Organization

Act in 1980. The mission of the agency is to:

- Strengthen the Federal commitment to assuring access to equal educational opportunity for every individual;
- Supplement and complement the efforts of states, the local school systems and other instrumentalities of the states, the private sector, public and private nonprofit educational research institutions, community-based organizations, parents, and students to improve the quality of education;
- Encourage the increased involvement of the public, parents, and students in Federal education programs;

- Promote improvements in the quality and usefulness of education through Federally supported research, evaluation, and sharing of information;
- Improve the coordination of Federal education programs;
- Improve the management of Federal education activities; and
- Increase the accountability of Federal education programs to the president, the congress, and the public. (Education.gov)

What is obvious from this lengthy mission statement is that the Department of Education has many tasks and serves many constituencies. With such a broad mission, or set of missions rather, it is unlikely that the agency is able to develop a strong sense of mission. This is one explanation for the lower performance on PART. Looking now to the key independent variables from chapter 3, the Department of Education is moderate (0.05), has a low leadership score (30 percent), and was created during divided government. All of these factors are associated with lower performance. Most strikingly, the Department of Education has a ratio of 3.47 appointees to every career staff person, which is by far the highest percent appointed ratio in the federal government. This alone predicts poor performance, as Lewis (2008) found and I confirmed in analyses of PART. In addition, the Department of Education is a redistributive agency and, as discussed earlier in this chapter, redistributive agencies fared the worst on PART. What role does culture play? The agency clearly has problems with goal ambiguity given its broad mission, which is reflected in its ranking on the FHCS performance-oriented culture index. The agency received a rating of 51 percent, which places it in the bottom third of agencies. Culture could certainly play a role in the agency's poor performance on PART, but there are many other factors at work too.

Using Clarify, the predicted agency PART score is 0.52 (95% CI = $0.49 <= \mu_1 - \mu_2 <= 0.54$) while the actual score is 0.50. The difference is minimal, and the average PART score falls within the confidence interval.

114

The Effect of Culture on Performance

I have discussed how culture, derived from the leadership, goals and mission of the agency, might affect agency performance. Each of the agencies discussed in this chapter have struggled to develop a performance culture for one reason or another and each performed differently on PART. There are many factors that influence the development of an agency's culture and, as I acknowledged in Chapter 3, one limitation of my analysis is that my "internal" variables might not be truly independent. Instead, these variables are likely influenced by one another and the broader agency context in which they operate.

To test the degree to which culture can be explained as a function of leadership, ideology and agency politicization, I ran a regression using my culture variable as the dependent variable. Interestingly, leadership, ideology and agency politicization (PERCENT APPOINTED), explain one-third of the variance in the culture variable (R-squared = 0.31). Leadership, as expected, is positively related to culture, which is statistically significant. Agency ideology is negatively and significantly related to culture, which suggests agencies with more conservative ideologies are less likely to have a performance-oriented culture. Politicization appeared to have no relationship with culture.

The role of culture is not well defined, but I am not alone in struggling to quantify this relationship. Apart from the difficulty in defining and operationalizing "culture," it may be that the relationship between culture and agency performance is indirect. Research on organizational culture in both the private and public sector suggests that communication could have a mediating or moderating effect on performance. Garnett et

al. (2008) synthesized decades of research on organizational communication, culture and performance and then empirically tested this multi-stage relationship between culture, communication and performance using state health and human services agencies. They found that there are certain types of communication present in organizations with a mission-oriented culture that increase performance. These forms include task-oriented feedback and upward communications. Despite their findings, however, the authors admit that a "pure type" of culture rarely exists. Usually organizations have a mixed culture, certainly true for federal agencies. These mixed or competing cultures, as Wilson (1989) and Ban (1995) refer to them, make it difficult for agencies to maintain a strong sense of mission.

While the literature and case studies offer some evidence that culture matters, much more empirical work needs to be done that offers a valid and reliable measure of culture and that explicates the direction and strength of the relationship between culture and performance. In my final chapter I review lessons learned from the empirical and exploratory analyses and discuss the present-day status of performance management in the Obama administration.

TABLE 4.1	Average Agency PART Scores	
Agency Name	**Average PART Score**	**No. of Programs Assessed**
Smithsonian Institution	.94	2
Federal Trade Commission	.92	1
Railroad Retirement Board	.91	2
Peace Corps	.91	1
National Science Foundation	.90	11
Trade and Development Agency	.90	1
Broadcasting Board of Governors	.88	8
Consumer Product Safety Commission	.88	1
Federal Deposit Insurance Corporation	.87	2
Nuclear Regulatory Commission	.87	7
African Development Foundation	.86	1
Tennessee Valley Authority	.86	3
National Credit Union Administration	.84	3
Department of State	.79	51
General Services Administration	.78	16
Securities and Exchange Commission	.76	5
National Aeronautics and Space Administration	.76	13
Social Security Administration	.76	3
Neighborhood Reinvestment Corporation	.75	1
Department of Defense	.75	54
Department of Energy	.75	55
Department of Transportation	.75	36
Department of Commerce	.73	33
Small Business Administration	.73	10
International Assistance Programs	.72	14
Department of the Treasury	.71	36
Appalachian Regional Commission	.67	1
Department of Homeland Security	.66	61
District of Columbia	.66	2
Department of Justice	.66	35
Army Corps of Engineers-Civil Works	.66	13
Department of Health and Human Services	.65	113

TABLE 4.1 Cont.	Average Agency PART Scores	
Agency Name	Average PART Score	No. of Programs Assessed
Department of Agriculture	.65	85
Court Services and Offender Supervision Agency for the District	.65	2
Millennium Challenge Corporation	.65	1
United States Interagency Council on Homelessness	.64	1
Department of the Interior	.64	70
Office of Personnel Management	.62	9
Environmental Protection Agency	.62	53
Department of Housing and Urban Development	.61	33
Department of Veterans Affairs	.60	10
Department of Labor	.58	35
National Archives and Records Administration	.53	2
Delta Regional Authority	.52	1
Corporation for National and Community Service	.51	4
Department of Education	.50	93
Federal Mediation and Conciliation Service	.50	1
Office of National Drug Control Policy	.47	5
Federal Communications Commission	.35	7

Table 4.2 Coding of PART Program Types by Lowi Typology	
Program Type	Policy Typology
Constituent	Capital assets and service acquisition programs
Distributive	Direct federal programs Research and development programs Competitive grant programs
Regulatory	Regulatory-based programs
Redistributive	Block/Formula grant programs Credit programs

Source: Greitens and Joaquin. 2010.

TABLE 4.3	Average Agency PART Scores by Type	
Agency Type	**Agency Name**	**Average Agency Score (Percentage out of 100)**
Regulatory	Department of Commerce Department of Energy Department of Justice Department of Labor Department of the Treasury	
Avg. PART Score for Regulatory Agencies		**69**
Distributive	Department of Agriculture Department of the Interior Department of Transportation Army Corps of Engineers	
Avg. PART Score for Distributive Agencies		**67**
Redistributive	Department of Education Department of Health and Human Services Department of Housing and Urban Development Environmental Protection Agency Small Business Administration Social Security Administration	
Avg. PART Score for Redistributive Agencies		**60**
Constituent	Department of Defense Department of Homeland Security Department of State Department of Veterans Affairs General Services Administration National Aeronautics and Space Administration National Science Foundation Office of Personnel Management USAID Smithsonian Institution	
Avg. PART Score for Constituent Agencies		**74**

Source: ExpectMore.gov

Table 4.4 Values of Key Independent Variables for Constituent Agencies

Agency Name	Agency Ideology	Leader1	Unified Govt (1 = yes, 0 = no)	Percent Appointed
Department of Defense	0.43	0.40	0	0.04
Department of Homeland Security	0.32	0.29	0	0.13
Department of State	-0.24	0.42	1	1.09
Department of Veterans Affairs	0.16	0.39	0	0.02
General Services Administration	0.20	0.46	1	0.31
NASA	0.22	0.46	0	0.08
National Science Foundation	-0.58	0.47	1	N/A
Office of Personnel Management	N/A	0.37	1	0.86
USAID	N/A	N/A	1	N/A
Smithsonian	N/A	N/A	1	N/A
Average Score	**0.07**	**0.41**		**0.36**

Table 4.5 Values of Key Independent Variables for Regulatory Agencies

Agency Name	Agency Ideology	Leader1	Unified Govt (1 = yes, 0 = no)	Percent Appointed
Department of Commerce	-0.13	0.37	1	0.46
Department of Energy	0.1	0.38	1	0.74
Department of Justice	0.19	0.37	1	0.34
Department of Labor	-0.16	0.36	1	0.93
Department of the Treasury	-0.27	0.35	1	0.08

Average Score	**0.13**	**0.37**	**N/A**	**0.51**

Table 4.6 Values of Key Independent Variables for Distributive Agencies				
Agency Name	**Agency Ideology**	**Leader1**	**Unified Govt (1 = yes, 0 = no)**	**Percent Appointed**
Department of Agriculture	0.16	0.34	1	0.21
Department of the Interior	-0.2	0.31	0	0.11
Department of Transportation	0.08	0.28	1	0.16
Army Corps of Engineers	N/A	0.39	1	N/A
Average Score	**0.01**	**0.33**	**N/A**	**0.16**

Table 4.7 Values of Key Independent Variables for Redistributive Agencies				
Agency Name	**Agency Ideology**	**Leader1**	**Unified Govt (1 = yes, 0 = no)**	**Percent Appointed**
Department of Education	0.05	0.31	0	3.47
Department of Health and Human Services	-0.42	0.39	1	0.21
Department of Housing and Urban Development	-0.02	0.32	1	0.87
Environmental Protection Agency	-0.5	0.34	0	0.36
Small Business Administration	N/A	0.26	1	1.35
Social Security Administration	-0.13	0.44	1	0.03
Average Score	**-0.11**	**0.34**	**N/A**	**1.0**

TABLE 4.8 Agency Rankings by Performance-Oriented Culture Index Scores		
Agency Name	**OPM Federal Human Capital Survey (2008)**	**Best Places to Work Survey (2007)**
National Science Foundation	65.4752	64.3
Nuclear Regulatory Commission	64.9185	68.4
Federal Trade Commission	64.0424	64.5
National Aeronautics and Space Administration	63.4694	59.0
Court Services and Offender Supervision Agency for the District	60.9899	61.1
Department of Commerce	59.3869	60.0
Department of State	57.1972	59.0
International Assistance Programs	57.1972	56.6
National Credit Union Administration	56.8816	65.8
Environmental Protection Agency	56.8347	56.2
General Services Administration	56.0771	60.9
Office of Personnel Management	55.5406	56.6
Department of Labor	54.4824	58.7
Department of Health and Human Services	54.4538	56.2
Department of the Treasury	54.2682	57.1
Department of Energy	54.1513	56.0
Department of Defense--Military	53.8968	56.6
Railroad Retirement Board	53.7706	51.4
National Archives and Records Administration	53.5351	55.2
Department of Justice	53.2625	57.6
Social Security Administration	52.8403	60.1
Federal Communications Commission	52.7569	53.3
Small Business Administration	52.5730	46.3
Securities and Exchange Commission	52.0804	62.8
Department of Education	51.8016	52.3
Department of the Interior	51.4731	52.2
Department of Agriculture	51.3430	53.9
Department of Veterans Affairs	51.1404	58.6
Department of Housing and Urban Development	49.0998	49.7
Department of Homeland Security	46.4461	47.8

Department of Transportation	46.1659	49.3
Broadcasting Board of Governors	40.9753	41.6

Sources: 2008 Federal Human Capital Survey, Office of Personnel Management; 2007, 2009 Best Places to Work Survey, Partnership for Public Service.

CHAPTER FIVE

Lessons Learned for Agency Performance and PART

Chapter Preview:

In this chapter I want to review what the empirical and exploratory analyses revealed about how external and internal agency characteristics affect performance. I then return to examine if this knowledge explains why the PART scores of the Basic Center and HOPE IV programs introduced in Chapter 1 are so different. In the last sections of the chapter I discuss the status of performance measurement and management in the Obama administration. I also provide suggestions for improving performance measurement and discuss directions for future research.

Empirical Findings Revisited

I set out to uncover why some agencies perform better than others and I believe I have several interesting answers. The Bush administration's PART scores, while imperfect, provide a measure of performance at the agency program level. In all, more than 1,000 agency programs received scores. Drawbacks of these scores include inconsistent grading and subjectivity on behalf of OMB graders. These concerns notwithstanding, I believe PART scores can reveal many things about why some agencies perform well.

In the first two chapters I discussed the framework for my analysis. The principal-agent theory attributes agency behavior to the external constraints imposed by the president and congress, who want to maintain agency accountability despite delegating policymaking authority to agencies. An alternate theory is that agencies are internally constrained by values, culture and other internally-derived factors that influence behavior and limit the extent to which the president and congress have control.

I developed a set of measures for both external and internal characteristics of agency behavior and then ran a series of regressions to determine how well these variables (and theories) explain performance. I found several variables met my expectations, which were based on previous studies of performance. The most consistent variables were PERCENT APPOINTED and LEADER1. Lewis (2008) found that more politicized agencies performed poorly compared to less politicized agencies. Politicization is measured by the ratio of appointees to career staff within an agency. LEADER1 is a measure of perceived leadership effectiveness from the Federal Human Capital Survey (FHCS). Measures of perceived leadership and culture have been validated in many previous studies (Brewer 2005; Wall et al. 2004; Bommer et al. 1995), although I must mention as a potential limitation the low response rate of the FHCS.

Other variables such as agency ideology and budget showed modest affects on performance. I also ran a special analysis of programs which received a Results not Demonstrated (RND) rating on PART. RND was given to programs that either failed to establish adequate outcome measures or collect suitable data to support those measures regardless of the overall numerical score. Because of the subjectivity OMB was afforded in making these determinations, I performed a Heckman selection model to test whether

some of my independent variables explained which programs received this rating. As I suspected, politics plays a role. Agency politicization (measured by PERCENT APPOINTED) is a highly significant predictor that agencies will receive a RND rating. This is evident for the Department of Education, which has the highest PERCENT APPOINTED ratio: 3:47 to 1, and the most programs rated RND (44 out of 93).

The two explanations of agency performance used to frame my analysis are not mutually exclusive. Agencies behave according to both external and internal factors. Some have negative influences on performance and others positively influence performance. My findings show that all are important to provide a complete explanation of agency performance.

HOPE IV and the Basic Center Programs Revisited

In my introduction to this dissertation I provided examples of two agency programs that despite many similarities such as budget size, constituencies served and program type, had very different performance scores on PART. I now return to these two programs to uncover whether the key variables found to significantly affect performance help explain the differences in these two programs.

The HOPE IV Severely Distressed Public Housing program is includedwithin the Department of Housing and Urban Development (HUD). The program was rated by PART in 2003 and received a numerical score of 42% and a rating of "inadequate." The Basic Center program for Runaway and Homeless Youth is housed in the Department of Health and Human Services (HHS), was rated by PART in 2006, and received a PART score of 85 percent and an "Effective" rating. Both are competitive grant programs. The programs' budgets for FY 2008 were 100 million (HOPE IV) and 96 million (Basic

Center). With so many similarities, however, what accounts for the vastly different

PART scores?

The key independent variables found to have a substantive and significant affect

on performance were the PERCENT APPOINTED, LEADER1, BUDGET, and

AGENCY IDEOLOGY, and PROGRAM TYPE. The budgets are similar and the

programs are the same type. But HHS, which houses the Basic Center program, has a

nine percent higher leadership score (39 to 31 percent) and a much lower percent

appointed score (0.21 compared to 0.87) than HUD, which houses HOPE IV.

Interestingly, HHS is a more liberal agency (-0.42 compared to -0.02), which does not

seem to affect the Basic Center program's success.

In Chapter 4 I provided a detailed discussion and exploratory analysis of agency

type and culture as predictors of performance. Both are redistributive agencies, which I

found to be disadvantageous for performance on PART. Reviewing the two agencies'

culture scores on the Federal Human Capital Survey (FHCS), it appears that HHS has a

higher score, 54 percent, compared to HUD's 46 percent.

The bottom line is that many factors contribute to performance on PART. Using

these two particular programs, it appears that percent appointed (or agency

politicization), leadership, and culture matter most. This is not true for all programs, but

these examples definitely highlight the need to account for both external (politicization,

budget, design) and internal (leadership, culture) performance factors.

Agency Performance and the Obama Administration

Even before assuming office, President Obama solicited the advice of top

performance management experts to develop a new and improved performance

management framework. According to Donald Kettl, while Obama did not immediately announce a new performance improvement program like Clinton and Bush, there is a "stealth revolution" underway to reshape government and performance (Kamensky 2010). These initiatives are part of three "mutually reinforcing performance management strategies" highlighted in the FY 2011 budget (OMB 2010). The first strategy is to use performance information to improve outcomes by setting high performance goals for agencies. The second strategy is to communicate performance information for better results and transparency. This includes developing online dashboards to keep the public apprised of agency progress. The third strategy is to strengthen problem-solving networks that go beyond the traditional performance model, centered on agencies and programs (Kamensky 2010; OMB 2010).

In the summer of 2009, agencies were asked to identify a small number of high-priority goals and the strategies for achieving them (Newcomer 2010). According to the High-Priority Performance Goal Initiative, these goals must demonstrate a high and direct value to the public or achievement of key agency missions rather than internal management or administrative issues; address how congressional authorizations and appropriations are required for implementation; identify how resolution of inter-agency collaboration and coordination challenges can improve effectiveness; set performance outcomes that can be evaluated, quantified and measured in a reasonable timeframe; and identify challenges that require a "concerted focus of agency resources" (OMB 2010; Orszag 2009). In addition, agencies are asked to define the problems and goals to be accomplished within the next 12 to 24 months. They must identify internal and external agency programs that can contribute to achieving the stated goals. Agencies must also

identify the lead staff to be held accountable for progress in achieving the goal(s) and outlining the strategy and key measures to be used to track progress (Orszag 2009). An example of one such goal is to "assist 3 million homeowners who are at risk of losing their homes due to foreclosure" (OMB 2010).

To address the second strategy, the Obama administration is taking several steps to improve the communication of performance information. Performance measurements and documents found not to be useful will be eliminated. Goals will "concisely and coherently" communicate what government is trying to accomplish. Performance measures, including agency, cross-agency and program-level, as well as those developed under GPRA and PART that proved useful to agencies, the public and OMB, will clearly state how well government is accomplishing goals (OMB 2010). Agencies will also be responsible for communicating progress toward goals to key audiences including the administration, congress, stakeholders and the public. To expedite this process, the OMB is developing a federal performance portal based on the Canadian system that will house federal goals and performance measures categorized by theme, agency, program and program type (Miller 2010; OMB 2010). The portal will offer links to mission-support management dashboards. Examples include the IT dashboard and the hiring dashboard that will provide information to all involved with these "common government functions."

The third strategy is to strengthen existing problem-solving networks and create new practitioner networks to encourage cross-agency collaboration on shared issues. Performance Improvement Officers appointed from every federal agency make up the Performance Improvement Council (PIC). The PIC, along with the OMB, is charged with developing a set of federal performance management principles, and a government-

wide performance management and implementation plan (OMB 2010). The PIC will also be the home of the federal communities of practice which provide tools and expert advice and are organized around problems, program type, and methods like quality management. OMB will meet with PICs from each agency to determine how well agencies are meeting their high-priority goals.

A parallel development is the recent passage (June 17, 2010) in the House of the "Government Efficiency, Effectiveness, and Performance Improvement Act of 2010" (H.R. 2142), which requires "quarterly performance assessments of Government programs for purposes of assessing agency performance and improvement," and the establishment of "agency performance improvement officers and the Performance Improvement Council" (OpenCongress.org). The bill, which now moves to the Senate, complements GPRA, which already requires federal agencies to define their missions and evaluate their performance.

The bill specifically requires federal agencies to: 1) identify and assess their goals; 2) submit strategic plans; 3) designate agency improvement officers within each agency; 4) establish a performance improvement council; 5) impose additional reporting requirements for each agency's annual plan; and 6) require several reports to the congress from the Government Accountability Office regarding the evaluation of federal programs. The Performance Improvement Council and Performance Improvement Officers were originally established in President Bush's Executive Order 13540, "Improving Government Program Performance." HR 2142 legislates their establishment as part of Obama's performance management and improvement efforts.

The president's FY 2011 budget requests $100 million for 17 agencies to conduct evaluations of their performance similar to those that would be required under H.R. 2142. The actual cost of the legislation depends on the extent to which agencies have integrated performance evaluation processes. The CBO estimates that completing the required program evaluation for all major federal agencies would cost about $150 million over the next five years (CBO 2010).

The Obama administration has announced a performance management "program," but the key question is how this approach differs from those of previous administrations. According to the FY 2011 budget, "the ultimate test of an effective performance management system is whether it is used, not the number of goals and measures produced" (OMB 2010). While GPRA and PART increased the number of agencies with performance measures, there is little evidence these measures are being utilized. Congress does not consider these measures while conducting oversight and there is little evidence that agencies are using them to improve performance (NAPA and GWU 2008). The other main problem with GPRA and PART is that the length of their review cycles prohibits short-term feedback to help agencies make immediate improvements. GPRA reports annually, while PART had a five-year review cycle, so neither could inform ongoing agency action.

How well Obama's program works remains to be seen, but the success of these new efforts depends on a number of factors. A very important component of a successful performance management effort is effective leadership within the top levels of the administration. Current OMB Director Peter Orszag has demonstrated his commitement to improving performance in a series of communications to agencies and the public.

Obama's appointment of Jeffrey Zients as the first-ever Chief Performance Officer is another indication the administration is taking these efforts seriously and committing tremendous resources to their success. In addition, the Performance Improvement Officers within each agency will be in place to utilize performance information to improve performance (Newcomer 2010a).

A second factor necessary for the success of Obama's performance management efforts is prioritization, which means focusing on a limited set of goals to be accomplished within the next one to two years. As mentioned above, this process is underway as all agencies developed high-priority goals by October 2009 (Newcomer 2010a).

The third, and perhaps most important, factor contributing to the success of the performance management efforts is the focus on use of performance data to improve programs. While GPRA and PART produced a wealth of performance data, there is little evidence these data have been put to good use. In a 2009 report for the IBM Center for the Business of Government, Shelley Metzenbaum outlined more than 20 recommendations to improve the use of performance data. Metzenbaum, now the associate director for performance and personnel management, suggested such steps as having focused, goal-driven meetings like the CompStat meetings in New York City, which successfully reduced crime. At these meetings, managers are expected," to bring evidence to discussions of problems and consideration of the options for addressing them" (Metzenbaum 2009). Some agencies have begun to adopt this approach.

It appears that the Obama administration is ambitiously working to improve performance management in the federal government and has sought the advice of leading

experts. There will, however, be challenges for the administration, especially

transforming management cultures within agencies toward more efficient use of

performance data (Newcomer 2010a).

Improving Performance Management

Throughout this dissertation I have mentioned PART's shortcomings including

measurement validity and reliability. An obvious question then is how to improve

performance measurement moving forward. Improving the reliability and validity of

performance measures is not an easy feat. One suggestion is to move away from using

questionnaires like PART which use "no" as the default answer. Another is to have

greater collaboration between those who design performance assessments and evaluations

and agency leadership. I believe a one-size-fits-all assessment or evaluation will

inevitably disadvantage some agencies.

Critics of PART suggested that it failed to take into account the different

environments in which programs and their respective parent agencies operate. From

what I have learned about PART, this is a valid criticism. Moving forward, performance

management must incorporate the elements of the agencies' task environments as

described in Chapter 4. It is impossible to determine how well agencies perform without

understanding the broader context and how it affects the job or jobs they have to carry

out. I tried to include measures of the agencies' tasks by controlling for type and the

broader environment by incorporating measures of constituency support.

As I mentioned in the previous section of this chapter, Shelley Metzenbaum, who

is currently serving in the Obama administration at the OMB, developed a list of

suggestions for how to improve performance management (Metzenbaum 2009). Perhaps

the most important suggestion is to ensure that the performance information generated is actually used to improve performance. There is evidence that PART scores were not used to improve agency performance or inform Congress' appropriations decisions. Even if performance data continues to be generated it will be for not unless it is used to improve future performance of federal agencies and their programs.

Conclusions and Future Research

The aim of my dissertation is to address why some agencies perform better than others. To frame my research I drew upon two broad and contrasting theories of agency behavior. The first, generally referred to as principal-agent theory, claims external actors (presidents and congress) use control mechanisms to influence agency behavior. These control mechanisms include appointments, oversight and budgeting. Studies, however, have found these influences sometimes have negative consequences for performance (Lewis 2003, 2008). In contrast to principal-agent theory, internal agency characteristics such as leadership and culture are offered as important predictors of performance. These characteristics or resources help agencies withstand external pressures and to exert their own influence on performance.

To answer my research question I used Program Assessment Rating Tool (PART) scores from the Bush administration. Scores for more than 1,000 agency programs provided a rich data source for my analysis. Potential limitations of this analysis, as discussed in detail in Chapter 3, include measurement validity and appropriate level of analysis.

I found through a series of regression analyses that both external and internal characteristics influence agency performance. The strongest, most significant predictors

of performance are agency politicization (PERCENT APPOINTED) and leadership. Future analyses should strive for more reliable and valid measures of internal agency characteristics, especially culture. In my qualitative analysis I tried to highlight culture's potential contribution to performance and I believe this is a rich area for further study.

Performance management remains especially important in today's political environment. Like many of his predecessors, President Obama has developed his own program to evaluate government performance. Due to increasingly advanced technologies, performance management information is likely to become more accessible and transparent which should make for higher quality analyses of the success of efforts to improve government performance.

REFERENCES

Balla, Steven J. 1998. "Administrative Procedures and Political Control of the Bureaucracy." *American Political Science Review* 92(3): 663-74.

Balla, Steven J. and William T. Gormley, Jr. 2003. *Bureaucracy and Democracy: Accountability and Performance*. Washington, D.C.: CQ Press.

Balla, Steven J., and John R. Wright. 2001. "Interest Groups, Advisory Committees and Congressional Control of the Bureaucracy." *American Journal of Political Science* 45(4): 799-813.

Ballenstedt, Brittany R. 2007. "NASA Urged to Align Internal, External Workforces." *Government Executive*. April 3rd. Available at: http://www.govexec.com/dailyfed/0407/040307b1.htm.

Ban, Carolyn. 1995. *How Do Public Managers Manage? Bureaucratic Constraints, Organizational Culture, and the Potential for Reform*. San Francisco: Jossey Bass Publishers.

Barnow, Burt. 2000. "Exploring the Relationship between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act." *Journal of Policy Analysis and Management* 19(1): 118–141.

Barrett, Katherine, and Richard Greene. 1999. *Grading the States 1999: A Management Report Card*. Washington, DC: Congressional Quarterly, Inc.

Baum, Lawrence. 1997. *The Puzzle of Judicial Behavior*. Ann Arbor: University of Michigan Press.

Bawn, Kathleen. 1995. "Political Control Versus Expertise: Congressional Choice about Administrative Procedures." *American Political Science Review* 89(1): 62-73.

Bertelli, Anthony., Clinton, Joshua., Grose, Christian., Lewis, David. and Nixon, David. 2008. "The Ideology of Federal Executives and their Agencies" *Paper presented at the annual meeting of the MPSA Annual National Conference, Palmer House Hotel, Chicago, IL.*

Brehm, John, and Scott Gates. 1997. *Working, Shirking and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor: University of Michigan Press.

Breul, Jonathan D. 2007. *Performance-Based Budgeting in the United States*. Washington, DC: IBM Center for the Business of Government.

Carpenter, Daniel P. 1996. "Adaptive Signal Processing, Hierarchy, and Budgetary Control in Federal Regulation." *American Political Science Review* 90(2): 283-302.

Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy*. Princeton: Princeton University Press.

Congressional Budget Office. 2010. "H.R. 2142 Government Efficiency, Effectiveness, and Performance Improvement Act of 2010." *As ordered reported by the House Committee on Oversight and Government Reform on May 20, 2010.* Available at: http://www.cbo.gov/ftpdocs/115xx/doc11556/hr2142.pdf.

Cueller, Henry and Todd Platts. 2010. "Cuellar & Platts: It's Time to Make Government More Accountable." *Roll Call.* http://www.rollcall.com/issues/55_147/ma_congressional_relations/47317-1.html.

Dahl, Robert. 1947. "The Science of Public Administration: Three Problems." *Public Administration Review* 7(1): 1-11.

Dahl, Robert. 1970. *Preface to Democratic Theory*. Chicago: University of Chicago Press.

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

Epstein, David, and Sharyn O'Halloran. 1999. *Delegating Powers: A Transactions Cost Politics Approach to Policy Making under Separate Powers*. New York: Cambridge University Press.

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

Frisco, Velda and Odd J. Stalebrink. 2008. Congressional Use of the Program Assessment Rating Tool," *Journal of Public Budgeting & Finance*. pp. 1–22.

Furlong, Scott R. 1998. Political Influence on the Bureaucracy: The Bureaucracy Speaks." *Journal of Public Administration Research and Theory*. 8:39-66.

Gallup Poll. July 2009. http://www.gallup.com/poll/121886/CDC-Tops-Agency Ratings-Federal-Reserve-Board-Lowest.aspx#1.

Garnett, James L., Justin Marlowe, and Sanjay K. Pandey. 2008. "Penetrating the Performance Predicament: Communication as a Mediator or Moderator of Organizational Culture's Impact on Public Organizational Performance." *Public Administration Review* 68(2): 266-281.

Gilmour, John B., and David E. Lewis. 2005. "Assessing Performance Assessment for Budgeting." *Journal of Public Administration Research and Theory* 16:169-186.

Gilmour, John B. 2006. "Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrating Budget and Performance." IBM Center for the Business of Government.

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

Government Accountability Office (GAO). 2005. *Performance Budgeting: PART Focuses Attention on Program Performance, but More Can be Done to Engage Congress.* GAO-06-28. Washington, DC: U.S. Government Accountability Office.

Greitens, Thomas and M. E. Joaquin. 2008. "Determinants of Program Performance: Resultsfrom OMB's PART Analysis," presented at the *Midwest Political Science Association Annual Conference*, Chicago, Illinois, April 3-6, 2008.

Greitens, Thomas and M. E. Joaquin. 2010. "Policy Typology and Performance Measurement: Results from the Program Assessment Rating Tool (PART)." *Public Performance and Management Review* 33(4): 555-570.

Gruber, Amelia. 2003. "OMB Ratings Have Little Impact on Hill Budget Decisions." *Government Executive Magazine* 13 June. Available at: http://www.govexec.com/dailyfed/0603/0613031.htm.

Gueorguieva, Vassia, Jean Accius, Carmen Apaza, Lamar Bennett, Clinton Brownley, Shea Cronin, and Panote Preechyanud. 2009. "The Program Assessment Rating Tool and the Government Performance and Results Act." *Public Administration Review* 39(3): 225-245.

Hall, Richard L. 1996. *Participation in Congress.* New Haven: Yale University Press.

Harsell, Dana M. and Jones, Vernon D. , 2002. "Managing for Results: Implementation Challenges Faced in New Jersey Municipal Government" *Paper presented at the annual meeting of the American Political Science Association, Boston, Massachusettes.* Available at*: http://www.allacademic.com/meta/p66470_index.html.

Hill, Carolyn J., and Laurence E. Lynn, Jr. 2005. "Is Hierarchical Governance in Decline? Evidence from Empirical Research." *Journal of Public Administration Research and Theory*. 15(2): 173-196.

Joaquin, M. Ernita and Sung Min Park. 2009. "Exploring the Topography of Federal Agency Performance and Effectiveness," Paper presented at the *Midwest Political Science Association Annual Conference*, Chicago, Illinois, April 2, 2009. http://www.allacademic.com/meta/p360751_index.html.

Kamensky, John M. 2010. "Obama's Performance Revolution: Changing How Government Works." IBM Center for the Business of Government. Washington, DC.   http://www.businessofgovernment.org/article/obama%E2%80%99s-performance-revolution-changing-how-government-works-03-2010.

Kettl, Donald F. 1993. *Sharing Power: Public Governance and Private Markets*. Washington, DC: The Brookings Institution.

Knack, Stephen. 2002. "Social Capital and the Quality of Government: Evidence From the U.S. States," *American Journal of Political Science* 46(4): 772-85.

Krause, George A. 1996. "The Institutional Dynamics of Policy Administration: Bureaucratic Influence Over Securities Regulation." *American Journal of Political Science* 40(4):1083-1121.

Krause, George A., and James W. Douglas. 2005. "Institutional Design versus Reputational Effects on Bureaucratic Performance: Evidence from the U.S. Government Macroeconomic and Fiscal Projections." *Journal of Public Administration Research and Theory* 15(2): 281-306.

Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance.* Princeton, NJ: Princeton University Press.

Lewis, David. E. 2003. *Presidents and the Politics of Agency Design*. Stanford, CA: Stanford University Press.

Light, Paul C. 1995. *Thickening Government: Federal Hierarchy and the Diffusion of Accountability*. Washington, D.C.: Brookings Institution Press.

Lowi, Theodore. 1964. "American Business, Public Policy, Case-Studies, and Political Theory." *World Politics.* 16:677-715.

March, James G., and Johan P. Olsen. 1989. *Rediscovering Institutions: The Organizational Basis of Politics*. New York: The Free Press.

McCubbins, Matthew D., Roger G. Noll, and Barry Weingast. 1987. "Administrative Procedures as Instruments of Popular Control." *Journal of Law, Economics, and Organization* 3(2): 243-77.

McCubbins, Matthew D., Roger G. Noll, and Barry Weingast. 1989. "Structure and Process as Solutions to the Politician's Principal Agency Problem?" *Virginia Law Review* 74(2): 431-82.

Meier, Kenneth J., and Laurence J. O'Toole. 2006. *Bureaucracy in a Democratic State: A Governance Perspective*. Baltimore: The Johns Hopkins University Press.

Meier, Kenneth J. and Kevin B. Smith. 1994. "Say It Ain't So, Moe: Institutional Design, Policy Effectiveness, and Drug Policy." *Journal of Public Administration Research and Theory*. 4(4): 429-442.

Melkers, Julia E. and Katherine G. Willoughby. 1998. "The State of the States: Performance-Based Budgeting in 47 out of 50." *Public Administration Review*. 58(1):66-73.

Melkers, Julia E. and Katherine G. Willoughby. 2004. *Staying the Course: The Use of Performance Measurement in State Governments*. Washington, DC: IBM Center for the Business of Government.

Metzenbaum, Shelley H. 2006. *Performance Accountability: The Five Building Blocks and Six Essential Practices*. Washington, DC: IBM Center for the Business of Government.

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

Miller, Jason. 2010. "OMB Preparing Performance Management Dashboard." Federal News Radio. June 17th. Available at: http://www.federalnewsradio.com/index.php?nid=110&sid=1982598.

Moe, Terry M. 1982. "Regulatory Performance and Presidential Administration." *American Journal of Political Science* 26(4): 197-224.

Moe, Terry M. 1985. "Control and Feedback in Economic Regulation: The Case of the NLRB." *American Political Science Review* 79(4): 1094-1116.

Moe, Terry M. 1989. "The Politics of Bureaucratic Structure." In John E. Chubb and Paul E. Peterson, eds. *Can the Government Govern?* Washington, D.C.: Brookings Institution: 267-329.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform.* Washington, DC: Georgetown University Press.

Newcomer, Kathryn. 2006. "How Does Program Performance Assessment Affect Program Management in the Federal Government." *Public Performance and Management Review* 29(June).

Newcomer, Kathryn. 2010. "Putting Performance First-A New Performance Improvement and Analysis Framework." *Framing a Public Management Research Agenda*. Washington, DC: IBM Center for the Business of Government.

Newcomer, Kathryn. 2010a. "Performance Management in the Obama Administration." *A Presentation to the American Evaluation Association*. Atlanta, GA. June 16, 2010.

Office of Management and Budget. 2010. "Analytical Perspectives, Budget of the United States Government, Fiscal Year 2011." Chapter 7. http://www.whitehouse.gov/omb/budget/fy2011/assets/management.pdf.

Office of Management and Budget. 2009. "The FY 2008 Performance Report of the Federal Government." http://www.whitehouse.gov/omb/expectmore/2008Performance.pdf. Accessed May 15, 2010.

Office of Management and Budget. 2004. "Rating The Performance of Federal Programs." http://www.gpoaccess.gov/usbudget/fy04/pdf/budget/performance.pdf. Accessed June 16, 2010.

Orszag, Peter. June 11, 2009. "Memorandum for the Heads of Departments and Agencies." Executive Office of the President. Office of Management and Budget. http://www.whitehouse.gov/omb/assets/memoranda_fy2009/m09-20.pdf.

Osterweil, Leon J., Lynette I. Millett, and Joan D. Winston, *Eds.* 2007. *Social Security Administration Electronic Service Provision: A Strategic Assessment.* A Report by the Committee on the Social Security Administration's E-Government Strategy and Planning for the Future, National Research Council.   Washington, DC: National Academies Press.

Partnership for Public Service and Gallup Consulting. 2008. "In the Public We Trust: Renewing the Connection between the Federal Government and the Public." http://www.ourpublicservice.org/OPS/documents/InthePublicWeTrustNov24.pdf. Accessed May 15, 2010.

Peacock, Marcus. 2008. "Testimony of Marcus Peacock, Deputy Administrator, Environmental Protection Agency before the Senate Committee on Homeland Security and Governmental Affairs, Subcommittee on Federal Financial

Management, Government Information, Federal Services, and International Security. http://hsgac.senate.gov/public/_files/PeacockTestimony.pdf.

Potosky, Matthew. 1999. "Managing Uncertainty through Bureaucratic Design: Administrative Procedures and State Air Pollution Control Agencies." *Journal of Public Administration Research and Theory* 9(4): 623-39.

Putnam, Robert. 1993. "The prosperous community: social capital and public life." *American Prospect*. 4(13).

Radin, Beryl. 2000. *Beyond Machiavelli: Policy Analysis Comes of Age*. Washington, DC: Georgetown University Press.

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

Radin, Beryl. 1998. "Searching for Government Performance: The Government Performance and Results Act." *PS: Political Science and Politics* 31(3): 553 555.

Rice, T.W. (2001). "Social capital and government performance in Iowa communities." *Journal of Urban Affairs*, *23*(3-4), 375-389.

Scholz, John T., and B. Dan Wood. 1999. "Efficiency, Equity and Politics: Democratic Controls over the Tax Collector." *American Journal of Political Science* 43(4): 1166-88.

http://www.ssa.gov. Accessed May 2010.

Tavits, Margit. 2006. "Making Democracy Work More? Exploring the Linkage between Social Capital and Government Performance." *Political Research Quarterly*, 59(2): 211-225.

Thompson, James D. 1967. *Organizations in Action: Social Science Bases of Administrative Theory*. Transaction Publishers: New Brunswick.

Towns, Steve. Ed. 2009. "Beth Noveck: Government Transparency Must Deliver Results." *Government Technology*. May 11th. Available at: http://www.govtech.com/gt/667712.

Trahant, Bill. 2008. "Realizing a Performance Culture in Federal Agencies." *Public Manager* July 14. Available at: http://thepublicmanager.org/cs/blogs/featured/archive/2008/07/14/realizing-a performance-culture-in-federal-agencies.aspx.

Weingast, Barry, and Mark Moran. 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy* 91(5): 756-800.

Whitford, Andrew B. 2005. "The Pursuit of Political Control by Multiple Principals." *Journal of Politics*. 67(1):29-49.

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

Wood, B. Dan. 1988. "Principals, Bureaucrats, and Responsiveness in Clean Air Enforcements." *American Political Science Review* 82(1): 213-34.

Wood, B. Dan. 1990. "Does Politics Make a Difference at the EEOC?" *American Journal of Political Science* 34(2): 503-30.

Wood, B. Dan, and John Bohte. 2004. "Political Transaction Costs and the Politics of Administrative Design." *The Journal of Politics,* 66 (February): 176-202.

Wood, B. Dan, and Richard Waterman. 1991. "The Dynamics of Political Control of the Bureaucracy." *American Political Science Review* 85(3): 801-28.

Wood, B. Dan, and Richard Waterman. 1993. "The Dynamics of Political-Bureaucratic Adaptation." *American Journal of Political Science* 37(2): 497-528.

Wood B. Dan, and Richard Waterman. 1994. *Bureaucratic Dynamics: The Role of Bureaucracy in Democracy*. Boulder, CO: Westview Press.