Institutionalizing Performance Management: Lessons for Government Leaders from the Government Performance and Results Act Modernization Act of 2010

Edgar Marion Hollandsworth III

Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Public Administration and Public Affairs

Matthew Dull, Ph.D., Associate Professor, Chair
David M. Bredenkamp
Adrienne Edisis
Patrick S. Roberts
Michael J. Meese

December 16, 2021
Arlington, VA

Keywords: Bureaucracy, GPRAMA, Management, Reform, Performance, QCA

Institutionalizing Performance Management: Lessons for Government Leaders from the Government Performance and Results Act Modernization Act of 2010

Edgar M. Hollandsworth

ABSTRACT

One of the most important trends in public management in the United States in recent history has been the spread of goal setting and performance measurement (Rainey 2014, 149), both core elements of a reform movement to make government agencies more transparent, accountable, and results-oriented (Light 1997). In the U.S. Government, the reform was legislated by the Government Performance and Result Act (GPRA) in 1993 and the GPRA Modernization Act (GPRAMA) in 2010 and implemented through Office of Management and Budget (OMB) policy guidance. Agency responses have ranged from limited use of performance information to comply with the law and only within the headquarters, to a voluntary embrace of its spirit and legislative intent and the routine, purposeful use of performance information at all levels to improve organizational learning, management decisions, and performance. The latter represents full institutionalization of performance management reform.

Empirical research on institutional and organizational conditions that lead to variance in institutionalization of performance management reforms predominates in state (e.g. Bourdeaux and Chikoto 2008; Taylor 2011; Moynihan 2008), local (e.g. Moynihan and Hawes 2012; Moynihan and Pandey 2010; Yetano 2013) and international contexts (e.g. Ohemeng 2011; Pollitt 2006). For the U.S. Government context, this study seeks to integrate and build on prior research on the effects of institutional or organizational conditions on observable indicators of reform institutionalization in agencies. These indicators include setting goals and performance monitoring (e.g. Lee and Kim 2012), performance information use (e.g. Dull 2009; Moynihan

and Kroll 2016; Moynihan and Lavertu 2012) or collaboration (e.g. Choi and Moynihan 2019). Using variable-oriented research designs, none of these studies have been able to characterize causal pathways or account for all conditions simultaneously, thereby limiting their generalizability. Also, conjunctural causation, causal asymmetry, and equifinality are common patterns of causation in organizational settings and in management scholarship; research designs that fail to account for these limit themselves to simple correlations (Furnari et al. 2020; Marx, Cambre, and Rihoux 2013). This study combines correlational and multi-conditional configurational qualitative approaches to assess causality, an unconventional research strategy in the public management literature. The author has additionally been unable to identify a single study that isolated the effect of bureaucratic type (James Q. Wilson 1989) on federal agency institutionalization of performance management. This study addresses this shortfall as well.

To further develop this line of research, this study employs the model of institutionalization progression proposed by Tolbert and Zucker (1996) and uses data from a 2017 Government Accountability Office survey of federal managers, an original online interview of 20 federal managers in 12 agencies, and interviews with 6 senior performance system managers and 8 fellows of the National Academy of Public Administration, to conduct a meso-level[1] mixed methods analysis employing hierarchical linear modeling, case research, and fuzzy set Qualitative Comparative Analysis (FSQCA) to identify which configurations of conditions best explain GPRAMA institutionalization in federal agencies.

Among other findings, quantitative models indicated an organizational culture with norms of accountability and empowerment, and organizational capacity for performance

[1] This study adopts the definition of "meso level" suggested by Berg-Schlosser et al. (2009) as "the level of organizations, social networks, collective actors, etc." (4).

evaluation, were by far the strongest individual predictors of reform institutionalization. When a range of theoretical antecedents of institutionalization are combined in qualitative configurational models, the study finds that agencies with immature performance management systems must first build senior leader commitment and supportive attitudes of managers for implementing the reform, and then seek to clarify organizational goals, in order to foster the adoption of performance measures. Agencies that are further along in the process of institutionalization can further the adoption of performance measures by investing in capacity for measuring performance and cultivating a culture of accountability and empowerment. To foster greater use of performance information, the study's findings suggest federal leaders should invest in capacity for performance evaluation and strengthen the credibility of their commitment to performance management reform, but that transforming agency cultures to become more results-oriented, often promoted in the literature, may not be necessary to achieve near-term improvements. Using James Q. Wilson's (1989) typology of bureaucratic designs, the study finds propositions based on it offer analytical leverage to explain variance in patterns of institutionalization of performance management practices between the studied US federal agencies. However, an unexpected finding was that Craft-type agencies are especially likely to use performance information for management decisions.

The study contributes 1) U.S. federal context to empirical research on "bottom up" factors mediating "top-down" reform policy implementation, 2) methodological innovation to public management research by employing hierarchical linear modeling to isolate agency-level effects; 3) original use of Qualitative Comparative Analysis (QCA) to identify configurational pathways to reform policy institutionalization in the U.S. Government; and 4) a transparent technique for classifying federal agencies by Wilson bureaucratic type. The study also offers

useful knowledge to legislators, stakeholders, political appointees, and career federal managers to shape the Federal Performance Management Framework and craft tailored management strategies suitable for the characteristics of federal agencies.

Institutionalizing Performance Management: Lessons for Government Leaders from the Government Performance and Results Act Modernization Act of 2010

Edgar M. Hollandsworth

GENERAL AUDIENCE ABSTRACT

One of the most important trends in public management in the United States in recent history has been the spread of goal setting and performance measurement (Rainey 2014, 149), both core elements of a reform movement to make government agencies more transparent, accountable, and results-oriented (Light 1997). In the U.S. Government, the reform was legislated by the Government Performance and Result Act (GPRA) in 1993 and the GPRA Modernization Act (GPRAMA) in 2010, and implemented through OMB policy guidance. Some agencies have complied with the letter of the law and used performance information to satisfy minimum requirements of legal compliance and only within the headquarters. Others have voluntarily embraced the spirit and legislative intent of the reform and started to use performance information at all levels to improve organizational learning, management decisions, and performance. The latter represents what this study defines as full institutionalization.

Scholars have long investigated the reasons why agencies implement performance management reforms differently and why some establish new routines more fully and permanently than others do. Much of this research has been at the state (e.g. Bourdeaux and Chikoto 2008; Taylor 2011; Moynihan 2008) and local levels (e.g. Moynihan and Hawes 2012; Moynihan and Pandey 2010; Yetano 2013), or conducted in other countries (e.g. Ohemeng 2011; Pollitt 2006). For the U.S. Government context, this study seeks to integrate and build on past research on which institutional and organizational conditions influence forms and degree of institutionalization of performance management practices in agencies. Examples include setting

goals and monitoring performance (Lee and Kim 2012), using performance information for management decisions (Dull 2009; Moynihan and Kroll 2016; Moynihan and Lavertu 2012) or collaboration (Choi and Moynihan 2019) . This past research has laid a solid foundation for a new phase of research that goes beyond simple case descriptions and correlational studies of individual variables to identify complex causal pathways across agencies and how these may differ depending on the presence or absence of political forces, resource dependency, oversight, and a host of organizational conditions like goal clarity, senior leadership commitment, an empowered and accountable culture, employee training, and the capacity for evaluation of performance data. The study also takes on a longstanding challenge in the study of public organizations by including agency identity as an additional organizational condition to test theoretical predictions about the adoption of performance management by bureaucracies having differing abilities to observe their outcomes and outputs (Wilson 1989).

To further develop this line of research, the field needs a better understanding of what combinations of conditions lead to deep institutionalization of the performance management practices of adopting performance measures and using performance information for management decisions. To contribute this, this study employs both statistical modeling and systematic cross-case comparative analysis of data from a large federal manager survey dataset from 2017 provided by the U.S. Government Accountability Office, an original online survey interview of 20 federal managers in 12 agencies, and interviews with 6 senior performance system managers and 8 fellows of the National Academy of Public Administration.

Among other findings, quantitative models indicated an organizational culture with norms of accountability and empowerment, and organizational capacity for performance evaluation, were by far the strongest individual predictors of reform institutionalization. When a

range of theoretical antecedents of institutionalization are considered together in qualitative configurational models, the study finds that agencies with immature performance management systems must first build senior leader commitment and supportive attitudes of managers for implementing the reform, and then seek to clarify organizational goals, in order to foster the adoption of performance measures. Agencies that are further along in the process of institutionalization can further the adoption of performance measures by investing in capacity for measuring performance and cultivating a culture of accountability and empowerment. To foster greater use of performance information, the study finds federal leaders should invest in capacity for performance evaluation and strengthen the credibility of their commitment to performance management reform, but that transforming agency cultures to become more results-oriented is not necessary to achieve near-term improvements. Using James Q. Wilson's (1989) typology of bureaucratic designs, the study finds propositions based on it offer analytical leverage to explain variance in patterns of institutionalization of performance management practices between the studied US federal agencies. However, an unexpected finding was that Craft-type agencies are especially likely to use performance information for management decisions.

The study contributes 1) U.S. federal context to empirical research on factors affecting policy implementation in complex bureaucracies, 2) methodological innovation to public management research by employing an unconventional statistical technique to study how agency identity impacts on institutionalization; 3) original use of Qualitative Comparative Analysis (QCA), a rarely-employed technique in the public management literature, to systematically compare conditions between federal agencies as separate cases to spot causal patterns; and 4) a transparent technique for classifying federal agencies by Wilson bureaucratic type. The study also offers useful knowledge to legislators, stakeholders, political appointees, and career federal

managers to shape the Federal Performance Management Framework and craft tailored management strategies suitable for the characteristics of federal agencies.

DEDICATION

I gratefully dedicate this work to my father, Edgar Marion Hollandsworth, Junior, a native son of old Virginia heritage, a 1957 Virginia Tech electrical engineering graduate and a leader in the Corps of Cadets, a successful manager in his long and distinguished career, and the very epitome of a Southern gentleman. Dad was proud that I followed in his footsteps to became a Hokie. In his final days, far more concerned for my welfare than for his cancer, he encouraged me to continue and finish what I had started and earn this doctorate. This one's for you, dad.

ACKNOWLEDGEMENTS

First, I salute all members of my committee. Professor Matthew Dull, my chair, devoted countless hours to guiding and encouraging me at every step, especially the times when I felt most critical of my own work. As one of the leading scholars in the field of public sector performance management, his own research and expertise inspired my research. Professor Adrienne Edisis helped me interpret the findings from my quantitative analysis, and kept me focused on the true prize: actionable insights for public management practice. Professor Patrick Roberts had undeniable bona fides as a fellow "bureaucracy watcher" and offered many helpful suggestions on both survey methodology and ways to socialize my research with the community of federal performance management practitioners. Professor David Bredenkamp reminded me that bureaucracies are organizations, and that, to explain their behavior, I needed to step outside my public management field comfort zone and consider relevant theories of organization. This led me to re-engage with theories of neo-institutionalism, and to learn something new about the under-studied phenomenon of organizational identity and its effects on organizational behavior. Finally, Dr. Michael J. Meese, Brigadier General, US Army (retired), adjunct professor of International Security Studies at Georgetown and visiting distinguished research fellow at the National Defense University Institute for National Strategic Studies, helpfully recommended a critical element be added to the study's inferential model, and introduced me to leaders in the National Academy for Public Administration (NAPA) who, in turn, opened many doors for me to interview top experts in this field and gain access to senior officials in the Office of Management and Budget. Proof positive that networks matter!

The Virginia Tech graduate school and Center for Public Administration and Policy always stand ready to help students with their administrative needs, and I had many over the

years. Sincere thanks go out to Myriam Lechuga, Corinne Julien, and Elia Amegashie for helping me with my course registrations and several updates to my plan of study.

This dissertation represents the support of many others at Virginia Tech and in the extended U.S. Government performance management community of practice. Shane Bookhultz and Minh Le at the Statistical Applications and Innovations Group at Virginia Tech confirmed the basic validity of the methods I used to measure the quantitative models and made several helpful suggestions for improvement, such as the use of multiple imputation techniques, which I implemented. My fellow CPAP PhD students, and several faculty members, helped me refine the draft survey instrument over three rounds of review, and were an outstanding sounding board during my concentration and capstone lectures. I am deeply in debt to John Kamensky, a national authority on good government who, as a NAPA fellow, connected me with leading experts in NAPA for my research focus and contributed substantially to the history chapter from lessons from his long and distinguished professional career. In federal agencies, I am indebted to Adam Lipton in the Office of Management and Budget and Natasha Schaaf in the General Services Administration for their assistance in identifying study volunteers for me to recruit. Twenty Deputy Performance Improvement Officers and performance management staff members completed the survey, six completed interviews, and 8 NAPA fellows were interviewed and provided sometimes startling insights that are not published anywhere else. I believe their contributions strengthened the validity, content, credibility, and overall value of this dissertation.

Above all, I want to thank my family. Growing up, debates about politics, government policies, and philosophy were nightly fare around our dinner table. My parents taught me and my sisters that, to be persuasive, one needs to back up one's claims with evidence and explain one's assumptions. My adult sons Bryce and Kyle learned this as well growing up and, as iron

sharpens iron, their feedback helped me to clarify my arguments and make the study more relevant for practice. Finally, as the last shall be first, so will I honor and thank the love of my life, my wife of 34 years, Delaine. Her patience with the demands of my studies was stretched thin, but it never broke. Despite the demands on both of us, she was always supportive, forced me to think long and hard about some of my assumptions, found lots of typos and other issues in my class papers, and stood by me through every phase of this program over the last 6 years. She shares equally in any recognition this study may receive. All critiques belong to me alone.

Table of Contents

List of Figures

List of Tables

Abbreviations

AIC Average Inter-item Correlation

BPA Behavioral Public Administration

CFO Chief Financial Officer

COO Chief Operating Officer

CPO Chief Performance Officer

CRS Congressional Research Service

DHS Department of Homeland Security

DOC Department of Commerce

DOD Department of Defense

DOJ Department of Justice

DOL Department of Labor

DOT Department of Transportation

DPIO Deputy Performance Improvement Officer

EO Executive Order

FSQCA Fuzzy Set Qualitative Comparative Analysis

FY Fiscal Year

GAO U.S. Government Accountability Office

GPRA Government Performance and Results Act

GPRAMA Government Performance and Results Act Modernization Act

HUD Department of Housing and Urban Development

IAD Institutional Analysis and Development

IQR Inter-Quartile Range

MAR Missing at Random

MCAR Missing Completely at Random

MGF Multiple Governance Framework

NAPA National Academy of Public Administration

NASA National Aeronautics and Space Administration

NDAA National Defense Authorization Act

NPM New Public Management

NPR National Performance Review

NRC Nuclear Regulatory Commission

OECD Organization for Economic Co-operation and Development

OI Organizational Identity

OLS Ordinary Least Squares

OMB Office of Management and Budget

OPM Office of Personnel Management

PART Program Assessment and Rating Tool

PI Use of Performance Information for Management Decisions

PIC Performance Improvement Council

PIO Performance Improvement Officer

PM Adoption of Performance Measures

PMA President's Management Agenda

PMS Performance Management System

PSM Performance System Manager

QCA Qualitative Comparative Analysis

SBA Small Business Administration

STEM Science, Technology, Engineering, and Math

VA Department of Veterans Affairs

VIF Variable Inflation Factor

Chapter 1. Introduction

> America is a land of remarkable energy, vitality, and imagination. But the complexity of its policy strategies could cause the country to choke on its own ambitions by failing to build a government that can achieve them. As the future unfolds, it is clear that the country will need to move even faster, that its problems will be even more interconnected, and that the consequences for failing to deal nimbly with those problems will be ever more serious (Kettl 2009, 236).

The American federal government finds itself increasingly trapped between citizen expectations to address disruptive and complex technological, social, and environmental upheaval in the world and limitations on its potential ability to do so as a result of growing fiscal pressures from annual deficits and accumulated debt. Public confidence in government suffers when it fails to adequately deliver the products and services people need or hard-earned tax dollars are wasted. Government failures happen when the future comes at governments faster than the imaginations of public officials can anticipate threats, when government leaders lack a vision or the skills to guide planning for high performance, or when governments do not invest in the human, technical, and material capacities needed to enable organizational or program changes to achieve that vision. On the other hand, when they occur, government failures often have an unforeseen beneficial effect: they can spur public dialogue about government reform. As long as government disappoints its citizens, and the government is democratic in nature, there will always be public demands for reform. This study is about one such reform and the lessons we can learn from its implementation about wise governance and public management strategies.

Background

America has substantial experience with reforming its federal government. Paul Light (1997) uses the metaphor of tides to frame the four reform movements in U.S. history. The most recent and current "tide" of government reform, which he dubs "Liberation Management," aims to make government work better and cost less by mandating that government agencies establish

customer service standards, adopt performance goals, and evaluate organizational and programmatic outcomes to guide management decisions that improve organizational performance in the governance context of public trust and administrative decentralization (39). Rainey (2014) hails the adoption of public sector performance measurement as "…one of the strongest trends in public management in the past two decades" (149). Van Wart (2012) notes performance measurement is one component of a strategic management system, along with strategic planning, program evaluation, and performance budgeting in modern organizations (290). Bryson (2011) argues that effective performance measures "…must help inform and guide strategy implementation" (292). Moynihan (2008) proposes a definition of performance management that synthesizes these elements: "…performance management [is] a system that generates performance information through strategic planning and performance measurement routines and that connects the information to decision venues, where, ideally, the information influences a range of decisions" (5). Boyne (2010) adds a critical additional element to Moynihan's definition: ". . . performance management is intended to direct the attention of public officials towards organizational achievements rather than inputs or procedures" (209). Integrating these, this study defines public sector performance management as a public management tool to measure organizational processes, outputs, and outcomes against organizational goals in order to inform strategic planning, program evaluation, budgeting, and a wide range of operational decisions.

Elected leaders and citizens have demanded that public bureaucracies be accountable and responsive and deliver services with high quality to citizens (Yang 2011, 149). Associate Director of the U.S. General Accounting Office, Christopher Mihm, testified in 2001 that "…the cornerstone of Federal efforts to successfully meet current and emerging public demands is to

adopt a results orientation" (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001, 12). In the U.S. Government, the reform goal of transforming agencies to adopt a results orientation generated requirements to evaluate federal department and independent agency performance that are now deeply embedded in federal laws such as the Government Performance and Results Act (GPRA) Modernization Act (GPRAMA), Pub. L. 111-352, 124 Stat. 3866 (2010), a substantial expansion of the GPRA, Pub. L. 103-62 (1993) (U.S. Government Accountability Office 2011a). GPRAMA was the most recent piece of major legislation designed to make the U.S. Government more accountable and perform better at lower cost to the American taxpayer (U.S. Government Accountability Office 2011a).

The offspring of the New Public Management and Reinventing Government movements of the 1990s, as described more fully in chapter 3, GPRAMA echoes a founding and enduring aspiration of American reformers since the Progressive Era. The Progressives sought to continuously improve the economy, effectiveness, and efficiency of government at all levels (Wilson 1887, 197; Croly 1909, 414; Skowronek 1984, 182-183; Stivers 1995, 523-524). The Progressive movement, in turn, was strongly influenced by the social survey research tradition originating in settlement houses focused on social welfare advocacy that were later adopted by municipal research bureaus to advocate for improvements in government efficiency, and by reforms to the civil service, budgeting, and cost accounting methods (Stivers 1995, 525-526). Williams (2002) notes the New York Bureau of Municipal Research developed the first performance measures between 1906 and 1912 (457). Frederick W. Taylor's Scientific Management principles were also a major influence (H. G. Frederickson et al. 2016; Heinrich and Marschke 2010, 183), and included setting individual performance standards for time spent

on tasks and using performance measurement to inform managers' decisions (Schachter 2007, 801). While the technical methods of measuring the performance of government organizations have fundamentally changed since the dawn of the 20th century, arguably the core aspirations of the reformers have not, and the challenges of institutionalizing the performance reform agenda have only expanded with the increasing size and complexity of the U.S. Government in the past century.

Today, variance in federal agency progress in institutionalizing performance management practices supporting GPRAMA affords scholars of public management, policy implementation, and public organizations a rich opportunity to investigate which combinations of institutional and organizational conditions most facilitate and impede the institutionalization of performance management reform policy. As this study will show, there is ample evidence from history (chapter 3) and current research to demonstrate that this variance exists in the case of GPRAMA institutionalization, and to falsify or confirm previous theoretical claims about the effects of these conditions on agency institutionalization of performance management. For example, the Partnership for Public Service and Grant Thornton, LLP interviewed and surveyed over 50 federal performance managers and hosted three focus groups in 2013 and found, *inter alia*, 1) federal agencies had advanced to different stages of development in building a performance culture under GPRAMA, 2) performance data was not always available or even being collected, 3) use of performance information for budgeting and other decisions was very limited, 4) resources for performance measurement were generally scarce, 5) performance managers needed better skills in interpreting performance data and educating employees on performance information use, and 6) most of the institutionalization had thus far been focused only at the top levels of agencies (Grant Thornton, LLP 2013, 2). Responding to continuing lack of progress by

some federal agencies in implementing GPRAMA, the government's lead oversight organization for monitoring agency compliance, the Government Accountability Office (GAO), made 47 recommendations to OMB and 53 recommendations to agencies, but found that only 36 percent of those made to agencies had been implemented by September 2017 (U.S. Government Accountability Office 2017a, 103). Results are summarized in table 1.1.[2]

Table 1.1. Agency Implementation of GAO Recommendations for GPRAMA (2017)

Fiscal Year	Implemented	Not Implemented	Percent Implemented
2013	0	3	0%
2015	8	17	32%
2016	11	14	44%
Total	**19**	**34**	**36%**

Source of data: U.S. Government Accountability Office. 2017. "Managing for Results: Further Progress Made in Implementing the GPRA Modernization Act, but Additional Actions Needed to Address Pressing Governance Challenges." GAO-17-775, 103. https://www.gao.gov/products/GAO-17-775.

Empirical evidence of variance in GPRAMA-related processes and work routines across federal agencies, including the adoption of performance measures and the use of performance information for management decisions, suggests latent institutional and organizational conditions may be largely responsible for variance in institutionalization decisions by federal managers. These conditions likely mediate federal manager responses to incentives created by Congress, oversight entities, and political appointees, and generate local organizational incentives[3] for

[2] The GAO recommendations agencies had not implemented by 2017 fell into five broad categories: 1) major management challenges such as setting performance goals and measures, establishing milestones, planning remedial actions, and designating responsible officials, 2) greater transparency in public reporting of quality of performance information to measure agency progress towards goals, 3) conducting effective data-driven reviews (i.e. leadership, goals, strategies, and programs), 4) Improving customer service standards to include performance targets and goals, and 5) training for performance management leaders (U.S. Government Accountability Office 2017a, 114-118).

[3] As discussed in chapter two, individual financial incentives play an inconsequential role for federal manager decisions to institutionalize GPRAMA, but a variety of non-monetary incentives

federal managers to either buffer their agency from performance management reform, or to advocate for its institutionalization. Heinrich and Marschke (2010) propose that policymakers can strategically structure public manager incentives in a way that improves the effectiveness of performance management systems and their contributions to agency performance and public value, although experience with "pay for performance" approaches has shown the limits of trying to use financial incentives in the public sector (184). If so, presumably incentive structures emphasizing factors other than compensation can be designed to improve the *ex ante* institutionalization of those systems in agencies as well. By strategically managing institutional and organizational conditions, federal managers and political leaders, under favorable conditions including holding required authorities, may be able to generate incentive structures that foster the institutionalization of more effective performance management systems.

Under favorable institutional conditions, such as legislation that mandates performance management work routines, the use of performance information for decision making may increase (Moynihan and Kroll 2016, 320). A recent official study, however, suggests performance information use across the U.S. Government has plateaued or even declined since 1997 (U.S. Government Accountability Office 2018),[4] roughly the start of the performance management reform period emerging out of the era of Reinventing Government and New Public Management (NPM). It is conceivable that the implementation of GPRA and GPRAMA, after a quarter century of work to institutionalize them, may have extended its reach as far as it can go,

---

appear to be strong and are associated with institutional and organizational conditions in this study.

[4] It is important to note that the GAO assessment was based on data collected from different samples of federal managers participating in a series of official surveys using the same questions, and is not based on a consistent sample in a longitudinal survey. Changes in individual expectations of performance information use by different sample members over time is not controlled in the GAO survey and may account for some of the variance in the data.

far short of universal institutionalization and many reformers' hopes. The culmination of these performance management reforms may add to what Ashworth et al. (2013) refer to as "…the sedimentation of successive reform programmes (sic)…" (S12). This prediction may foretell yet another phase in the evolution of performance-related reforms that continues to build on what remains from the previous ones.

For public agencies, their politically appointed leaders, and career civil service managers, agency performance should theoretically matter greatly, not only out of a sense of public service motivation to create public value (Moore 1995) but also, from public choice and institutional theory perspectives, to support their agency's autonomy (Carpenter 2001; Curry et al. 2015; Nielsen 2014); reputation (Busuioc and Lodge 2016; 2017; H. Doering et al. 2021; Moynihan 2012; Carpenter and Krause 2012); and budget (Niskanen 1994; Ryu 2011). Public perceptions of government performance affect citizens' trust and satisfaction in their government (Beeri, Uster, and Vigoda-Gadot 2019; Kettl 2016) as well as their confidence in its economy, efficiency and effectiveness (U.S. General Accounting Office 2000, 15). If these views deteriorate too far, public faith in government could be put at risk (Kettl 2009, 28). A study by the Organisation for Economic Co-Operation and Development (1997) links performance with democratic legitimacy and confirms the international relevance of Kettl's concerns:

> In many OECD Member countries, there is currently an ideological and technical discussion on the legitimacy of government in society. Democratic deficits appeared and gaps between the state and its citizens became obvious to politicians and the general public. Service delivery by the public sector has become a key issue in reconstructing the legitimacy of government. (7)

The preponderance of the literature also shows a positive association between the institutionalization of performance management systems and actual agency performance in the public sector, especially when combined with sharing best practices (Boyne 2010; Gerrish 2016,

62-63). A theoretical claim could be made that, since there are possible tangible and psychological payoffs for doing so, many public managers should be self-interested in improving their agency's performance, even in the absence of mandating legislation. However, as chapter 3 will show, institutional and organizational conditions have often placed limits on their freedom of action to embed performance management regimes in their agencies.

The adoption of performance measures and the use of performance information for management decisions, two primary indicia of performance management system institutionalization in this study, nevertheless vary greatly between public organizations (de Lancer Julnes and Holzer 2001; Radin 2006). U.S. Government Cabinet-level departments and independent agencies demonstrate significant variance in degree of GPRAMA institutionalization (Government Accountability Office 2018). Based on the review of the literature in chapter 2, management decisions to institutionalize GPRAMA are likely influenced by federal managers' perceptions of the risks and incentives associated with embedding performance management practices.

Risk aversion and blame avoidance have been shown to be powerful incentives for public managers in the public management literature (e.g. Hood 2007; 2011; Mortensen 2013) as well as for political leaders legislating and overseeing reform implementation who may use performance measurement to shift blame to administrators (Dull 2006, 108; Nielsen and Baekgaard 2015). Hong, Kim, and Son (2020) observe that the blame avoidance literature has reached a consensus that blame avoidance is an even more powerful incentive than credit claiming for public managers (1243). This view conforms to empirical research on citizen evaluations of public sector performance information, which have revealed a negativity bias that leads to a no-win situation for public managers in which "bad performance is punished while

good performance is not rewarded" (Olsen 2017, 567). Vice President Gore noted in 1998 that ". . . creating honest measures will create risks" (Council for Excellence in Government and National Academy for Public Administration 1998). Kettl (2016) observes some public managers are leery of the risks they may run if they fully implement performance management legislative requirements:

> For many players in government's world. . .information is often viewed as a gotcha game, to identify problems, or as a defense against investigations, to minimize hassles. And that, in turn, makes it harder to convince all the players to use information to support positive steps like making programs work better, because many government insiders look on information with profound suspicion (155).

Benjamin and Posner (2018) further suggest public managers must consider conflicting values and potential political consequences as they exercise accountability (580). Clearly, it should surprise no one that convincing federal managers they or their agencies have much to gain and little to lose by institutionalizing performance measurement practices and reporting program performance would be a hard sell.

Institutional and organizational contexts generate both incentives and disincentives for public managers to participate (Simon 1997, 144-145) and innovate (Hopkins 2016, 343). Examples of the wide range of strong and weak incentives for public managers proposed in the literature are described in detail in chapter 2. Perceived risks and incentives matter for reform policy institutionalization because, as Rohr (2002) notes, public managers exercise administrative discretion in how they implement policies (135). Public sector research into the processes of organizational change has demonstrated a "critical" role for public managers (Fernandez and Rainey 2006, 168). Hatry et al. (2003) synthesize common claims in the public management literature about public managers' agency in a governance environment that constrains their authorities:

> Despite limited control over what they do, federal program managers can have considerable influence over [a] program's procedures. They have a role in developing the proposed budgets for their programs, and they are key to motivating their own personnel and contractors to perform. Although such actions are not nearly as "global" or influential as those of policy officials, they nevertheless can have important and significant effects on the public. (6)

Therefore, if institutional and organizational contexts are favorable and create enough positive incentives for public managers to perceive that the payoffs outweigh the risks to them and their agencies from implementing the reform, then it is more likely a public agency with such incentivized managers will implement performance management reform. The opposite should also be true.

Purpose of the Study

Given the problem of perceived risks in institutional and organizational contexts facing public managers who desire to institutionalize performance management processes in their agencies, discovering the relevant antecedents of variance in the acceptance, adoption, and routine use of performance management practices across agencies affords scholars of public management potential new insights into how these antecedents shape the incentives for public managers to embrace legislative reforms at the agency and whole-of-government levels. With a clear understanding of these factors and how they interact during the institutionalization of reform legislation inside federal agencies, legislators, political appointees, policymakers, and public leaders and managers can shape the governance environment and attributes of their organizations in a way that sets the stage for successful reforms and improved agency performance.

In light of these potential benefits for practice, interest in this research challenge among public management scholars is current and substantial, and there is ample room for growth in both theory development and empirical testing and validation. O'Toole and Meier (2015)

advocate bringing "…the systematic study of context into theories of public management research, and especially those theories seeking to explain the relationship between management and performance" (238). Nitzl, Sicilia, and Steccolini (2019) call for more research to explain under which conditions public organizations make use of performance management (687). Moynihan et al. (2011) advocate for a Minnowbrook III normative perspective that analysis of performance management regimes requires an appreciation for the governance context in which they must operate and advance an agenda for research that challenges scholars to investigate ". . .the likely response to the introduction of performance regimes—passive, political, purposeful, or perverse. . ." and ". . .factors [that] encourage different forms of performance information use" (i142-i144). According to Moynihan (2009), passive responses mean agencies merely comply with the law but do not try to use it for management decisions, purposeful responses mean taking steps to improve program performance using performance information, political responses are observed when public managers seek to curry favor with political leaders by reporting performance that casts a favorable light on their agencies, and perverse responses occur when managers manipulate measures to game the performance management system (593).

Of singular importance in the literature in the past 20 years towards these research aspirations has been the theoretical articulation of indicia of agency responsiveness to performance management reform mandates. Two prominent vectors of responsiveness to reform mandates often studied are public manager selection and adoption of performance measures (e.g., Andrews, Boyne, and Walker 2011; de Lancer Julnes and Holzer 2001; Frey, Homburg, and Osterloh 2013; Radin 2006) and their use of performance information for management decisions (e.g., Boyne 2010; Kroll 2015; Moynihan and Hawes 2012; Moynihan and Pandey 2010).

In the public management literature on performance management, discussed in detail in chapter 2, extended threads of theory and empirical research into external conditions of official oversight and political and resource dependency influences and seven common organizational conditions within government agencies have sought to characterize the relative level of effect each has on agency adoption of performance measures and the use of performance information at the state, local, and national levels. The organizational conditions include 1) goal clarity (or its inverse, goal ambiguity), 2) an innovative culture of accountability and empowerment, 3) the credible commitment of agency leadership, including management, 4) employee training in performance management, 5) the capacity for evaluation, including staff, technology, and business processes, 6) the maturity of performance measures, which is the level of routinized use of performance measures and quality and timeliness of performance information they generate, and 7) bureaucratic identity, based on the taxonomy of four bureaucratic types proposed by James Q. Wilson (1989), the latent performance orientation of an agency based on the observability of its outputs and outcomes.

Empirical research exploring the conditions associated with the variance in institutionalization of performance management-related reforms predominates in state (e.g. Bourdeaux and Chikoto 2008; Genest-Grégoire, Charbonneau, and Bromberg 2018; Taylor 2011; Moynihan 2008), local (e.g. Moynihan and Hawes 2012; Moynihan and Pandey 2010; Yetano 2013) and international contexts (e.g. Mauro, Cinquini, and Grossi 2018; Ohemeng 2011; Pollitt 2006). For the U.S. Government context, this study seeks to integrate and build on prior research on the effects of institutional or organizational conditions on observable indicators of reform institutionalization in agencies, such as setting goals and performance monitoring (J. W. Lee and Kim 2012), performance information use (Dull 2009; Moynihan and Kroll 2016;

Moynihan and Lavertu 2012) or collaboration (Choi and Moynihan 2019) . Using only variable-oriented research designs, none of these studies have been able to characterize causal pathways or account for all conditions simultaneously, thereby limiting their generalizability. Also, conjunctural causation, causal asymmetry, and equifinality are common patterns of causation in organizational settings and in management scholarship; research designs that fail to account for these limit themselves to simple variable correlations (Furnari et al. 2020; Marx, Cambre, and Rihoux 2013). Accounting for the equifinal nature of complex causation in organizational and strategic contexts can offer analytic traction to organizational theory scholars investigating alternative management strategies for organizational design intended to achieve an outcome (Fiss 2011, 394). The author has additionally been unable to identify a single study that empirically isolated the effect of bureaucratic type (James Q. Wilson 1989) on federal agency institutionalization of performance management reforms. The design of this study addresses these needs.

Models for analyzing the relationship between institutions and public manager decisions have been limited to only a few topic areas such as environmental and municipal governance, and there is little published research modeling institutional conditions and how public managers respond to them, a lacuna in the scholarship that has lead scholars to call for more empirical research using institutional analysis methods to study policy implementation, among other topics (Baldwin, Chen, and Cole 2019, 891, 895). Other scholars have advocated that research into organizational practices investigate the external and internal causes not only for the adoption of the reform, but also for its implementation in practice (MacIndoe and Barman 2013, 733). This research seeks to respond to these concerns by providing evidence and analysis for how

governance and organizational conditions affect performance management policy institutionalization in the context of U.S. Government agencies.

The purposes of this research, then, are both confirmatory and exploratory. The primary goal is to falsify or validate claims in the literature concerning the direct and combinatorial effects of external oversight and political influences, and the seven organizational conditions theorized in the public management literature described above, on the institutionalization of GPRAMA performance management reforms across the 24 U.S. Government departments and independent agencies covered by the 1990 Chief Financial Officers' Act (henceforth all are referred to simply as "agencies" for simplicity). A second goal is to test the predictions of James Q. Wilson concerning how the observability of organizational outputs and outcomes, which classify the type of bureaucracy an agency is, affects agency adoption and use of performance management practices. Understanding how deeply-institutionalized conditions like bureaucratic identity affect the ability of government agencies to integrate a robust regime of performance management may have a range of implications for the future development of performance management practices in public agencies. Understanding the relative strength of these effects, and how and why they vary between agencies, will provide valuable findings about not only the limits of performance management reform in specific agencies, but also a fuller understanding about conditions favorable to the successful implementation of government reforms more broadly across the U.S. and perhaps other large and complex national governments with similar governance and organizational structures.

Empirical findings may also have implications for the ongoing debate concerning the theoretical relevance and appropriateness of the New Public Management philosophy for the diverse and complex governance structures of the U.S. Government. These structures are based

on a constitutional order that is fragmented and inefficient *by design*. Specifically, the findings in this study on the limits of institutionalization of reform may be grist for the enduring normative debate in the public administration literature about the appropriate balance between the values of programmatic efficiency and effectiveness in the execution of policy and the values of democratic accountability, transparency, and citizen engagement in the federal American political system (e.g. D. G. Frederickson and Frederickson 2006; Green and Hubbell 1996; Moe 1994; Molina 2015; Radin 2006). Limits on GPRAMA institutionalization may ultimately be a function of systemic balancing of these competing values. This balancing transpires quietly behind the institutional veil. Its outcome serves as a kind of canary in the coal mine to warn public policymakers that performance reforms are worthwhile, but only to the point at which they begin to place the founders' vision of democratic accountability of federal agencies at risk.

Research Questions and Overview of Study

This study seeks to answer two primary research questions:

1) How do external oversight, political influence and resource dependency, as well as organizational conditions affect agency adoption of performance measures and use of performance information for management decisions?

2) How does the nature of an agency's identity according to its bureaucratic type, specifically Production, Procedural, Craft, or Coping (J. Q. Wilson 1989) affect its adoption of performance measures and use of performance information for management decisions?

These two broad questions are unpacked in chapter 2 with conceptual definitions and literature-informed hypotheses concerning the effects of contextual and organizational conditions on the observable routines of performance management. The presence of these routines serves as a dependent variable in the quantitative portion of the study and as a condition in the

qualitative section. An historical institutionalist account of modern performance management reform in the U.S. Government, federal agency responses to it, and the influence of contextual and organizational conditions in chapter 3 illustrates the definitions for these proposed in chapter 2. The history offers evidence of the extent of their effects on past agency choices concerning how and when to institutionalize reform practices. The empirical evidence in chapter 3 is enhanced by interview testimony of former federal executives who participated in the development and implementation of GPRA, PART and GPRAMA and even today continue to stay engaged in the community of good government through the National Academy of Public Administration (NAPA).

The operationalization of variables and conditions for quantitative and qualitative research models in this study, respectively, is introduced in chapter 2 and then further described in the research design sections in chapters 4 and 5. Chapter 4 is focused on research question one and employs a mixed methods approach that is sufficiently robust and efficient to generate at least modestly-generalizable and preliminary findings at agency, bureaucratic type, and whole of government levels of analysis. First, using a 2017 GAO survey of over 3,000 federal managers, the study employs cross-sectional exploratory and confirmatory factor analysis of the GAO survey dataset to generate latent macro-variables, and then uses these in multi-level mixed effects and generalized ordered logistic regressions to measure the correlations of external oversight and seven organizational antecedents with the institutionalization of performance management routines at the agency and whole of government levels. Based on claims of the GAO about its survey methods, the quantitative analysis assumes the sample of personnel from the 24 federal agencies in the analyzed dataset, which employ over 90 percent of all federal employees, offers a broad-based sample of federal agencies and the federal workforce to permit

the development of inferences about institutionalization outcomes and their antecedents across a significant subset of the government and within federal agencies.

To further validate and add context to the quantitative analysis of survey data, and to provide deeper qualitative knowledge about how agency institutionalization of GPRAMA is shaped by these same conditions as well as political influences, chapter 4 also employs fuzzy set QCA, a set-theoretic comparative case-based qualitative method, to analyze 2021 survey data collected from 20 federal performance system managers in 12 agencies to measure the effects of all theoretically-plausible causal combinations of environmental and organizational conditions in the studied federal agencies. The goal of this research is to identify the most probable combinations of conditions that are necessary or sufficient to enable agencies to achieve high levels of institutionalization of GPRAMA. The comparative portion aims to achieve modest generalizability of causal arguments based on systematic cross-case comparisons using well-established set-theoretic methods widely used in scholarly research in a variety of social science fields and disciplines. Empirical data to characterize the presence of causal conditions, performance system manager incentive structures and attitudes towards performance management, and where possible causal mechanisms within federal agencies, is collected via interviews with federal performance system managers. Responses are then organized by category of predictor in the inferential model, and compared to predicted causal configurations found in the QCA-based study using cross-case synthesis techniques in order to validate quantitative and QCA findings and provide micro-level data to support and illustrate research findings.

Chapter 5 seeks to answer research question 2 using a quantitative strategy similar to that used in the quantitative portion of chapter 4. Here, the study focuses on between-agency

comparisons of the random effects of agency identity and bureaucratic type as independent variables on the primary indicators of GPRAMA institutionalization in order to test theoretical propositions derived from James Q. Wilson's taxonomy of bureaucracies in his classic book *Bureaucracy* (1989). It also estimates within-agency fixed effects of bureaucratic identity and bureaucratic type on agency adoption of specific kinds of performance measures and specific uses of performance information in order to generate useful findings for public sector practice.

Chapter 6 concludes by summarizing and integrating the study's findings and evaluating their significance, noting their potential contribution to the literature and to public management practice. Substantive and methodological recommendations for further research are also presented.

Chapter 2. The Inferential Framework and Review of the Literature

Drawing from multiple strands of theory in the organizational, public management, and policy implementation literatures, this chapter sets the stage for the historical institutionalist and mixed-methods empirical analyses to follow. The chapter begins with a short introduction to relevant institutional and organization theories. It also presents an established model for measuring the progress of institutionalization of structures in organizations that is useful for this study's goal of evaluating the degree of adoption and routine use of performance management practices in government agencies. The chapter then reviews scholarship in the public management and performance management literatures regarding conditions which may affect institutionalization of performance management reforms and integrates this information into a high-level inferential model with multiple hypotheses to test in this study.

The discussion of the institutionalization of performance management in public agencies begins with grounding it on the foundation of relevant propositions from organizational and institutional theories along with supporting empirical findings drawn from related public management contexts. Somewhat surprisingly, the literature on organizational theory is poorly connected with the public management literature, and, despite some early usage, empirical organization research using public agencies has significantly declined (Pfeffer 2006, 458-459). Organizational theory scholarship has a lot to say, nevertheless, about the concept of institutions, how institutions interact with organizations, and how one can evaluate the diffusion and formalization of institutional values and rules in organizations. This is precisely the challenge faced by performance management reformers.

Public administration scholarship has been criticized for a perceived lack of scientific rigor due to insufficient use of theory and weak methodologies. Critics argue the field would

benefit by invoking models from organizational theory, perhaps qualifying such scholarship for publication in prestigious journals of organizational theory (Vogel 2014, 399). Organization science includes a universe of conceptual frameworks. One of the more influential is institutional theory (Fuenfschilling and Truffer 2014, 774). A study that employs institutional theory to analyze public management reforms, as offered here, should help to create a bridge between public management and organizational theory and help to re-establish a long-lost scholarly tradition.

Organizational Theories as a Foundation to Study Reform Policy Institutionalization

In colloquial speech, the concept of an "institution" is often conflated with a well-established organization. Based on an open systems perspective, it is more accurate to state that organizations, are ". . . systems of interdependent activities. . ." that may be coupled tightly or loosely and must be constantly ". . . produced and reproduced. . ." to survive (Scott and Davis 2007, 31). Instead of being institutions, organizations exist within institutional frameworks. One authoritative definition of institutions describes them as "…cultural-cognitive, normative, and regulative elements that, together with associated activities and resources, provide stability and meaning to social life" (Scott 2001, 48; as quoted in Scott and Davis 2007, 258). Barley and Tolbert (1997) insightfully observe that "…institutions are to social action as grammars are to speech" (96). With collective action contexts in mind, Ostrom (1990) defined institutions as "the sets of working rules that are used to determine who is eligible to make decisions in some arena, what actions are allowed or constrained, what aggregation rules will he used, what procedures must be followed, what information must or must not be provided, and what payoffs will be assigned to individuals dependent on their actions" (51).

Institutions affect organizations profoundly because there are potential payoffs for survival, legitimacy, and in some cases performance for organizations that integrate them either in daily practice, or at least symbolically. Meyer and Rowan (1977) claim "organizations are driven to incorporate the practices and procedures defined by prevailing rationalized concepts of organizational work and institutionalized in society. Organizations that do so increase their legitimacy and their survival prospects, independent of the immediate efficacy of the acquired practices and procedures" (340). Indeed, Abrahamson (1991) and Strang and Macy (2001) note that the diffusion of new practices may even reduce efficiency or harm an organization (as cited by Ansari, Fiss, and Zajac 2010, 70). A recent meta-analysis of the literature on organizational adaptation supported claims that adaptation fosters organizational convergence with institutional environments, but found effective adaptations do not necessarily predict organizational performance or even survival, while maladaptation may in some cases lead to improved performance due to effective organizational differentiation in competitive environments (Sarta, Durand, and Vergne 2021, 58, 60, 61). The relevance of the claim concerning exceptional cases in which maladaptation creates certain benefits may, however, be debated in monopolistic public sector contexts where there is little or no competition between government organizations, and maladaptation may risk harm to an agency's legitimacy or reputation.

A subset of institutional theory which appears particularly promising for exploring this study's research questions is "new" institutionalism, or neo-institutionalism.[5] Greenwood, Hinings, and Whetten (2014) observe that neo-institutionalism is an important component of both organization and management theories and argue in favor of making greater use of

[5] As compared to "old" institutionalism scholarship, exemplified by classic authors such as Selznick (1957) who focused on power, values, informal structures, etc. (Greenwood, Hinings, and Whetten 2014, 1207).

institutional theory to understand organizations (1207). Neo-institutionalism, in turn, borrows foundational concepts from "old institutionalism" such as power, structure, and organizational legitimacy and has three branches which are distinct, yet self-identify as neo-institutionalist: historical institutionalism, rational choice institutionalism, and sociological institutionalism; adding further complexity, each branch has adherents who subscribe to different assumptions about how individuals respond to institutional conditions, i.e. either rationally and strategically, or based on personal habits, mindsets, and cultural norms in their social environment (Hall and Taylor 1996, 936-939). Therefore, there is no single form of neo-institutionalism. Because this study interprets evidence through the lens of historical and sociological institutionalism in differences places, a brief description of each is in order.

Historical institutionalists define institutions as ". . . the formal or informal procedures , routines, norms and conventions embedded in the organizational structure of the polity or political economy" (Hall and Taylor 1996, 938) and emphasize that institutions are a critical factor that shapes the evolution of organizations and collective actions over time:

> . . . historical institutionalists. . . .have been strong proponents of an image of social causation that is 'path dependent' in the sense that it rejects the traditional postulate that the same operative force will generate the same results everywhere in favor of the view that the effects of such forces will be mediated by contextual features of a given situation often inherited from the past (Hall and Taylor 1996, 941).

Thus, established agency practices for performance measurement, for example, constitute the instantiation of the institution of performance management in an organization, as mediated by pre-existing contextual conditions. Historical institutionalism, therefore, offers a helpful lens to interpret the historical record of GPRAMA institutionalization across a variety of different agency and historical contexts, and is employed in chapter 3.

Sociological neo-institutionalists believe "…the social environment [affects] the identities, behaviors and practices of people and groups now conceived or constructed as bounded, purposive and organized actors" (Meyer 2017, 833). Canonical sociological neo-institutionalist studies postulate the homogenization of organizational structures over time due to the forces of institutional isomorphism (with its coercive, mimetic, and normative mechanisms), structuration in organizational fields, and organizational legitimization (e.g. DiMaggio and Powell 1983; 1991; Jepperson 2002; Tolbert and Zucker 1983):

> Once disparate organizations in the same line of business are structured into an actual field (as we argue, by competition, the state, or the professions), powerful forces emerge that lead them to become more similar to one another. . .Early adopters of organizational innovations are commonly driven by a desire to improve performance. . .As an innovation spreads, a threshold is reached beyond which adoption provides legitimacy rather than improves performance (Meyer and Rowan 1977). Strategies that are rational for individual organizations may not be rational if adopted by large numbers. Yet the very fact that they are normatively sanctioned increases the likelihood of their adoption" (DiMaggio and Powell 1991, 65).

A more recent extension of sociological neo-institutional research examines root causes for non-isomorphic patterns of structure and practices in organizations, and the interactions between external and organizational conditions that explain variance in the institutionalization of practices between organizations (e.g. Mauro, Cinquini, and Grossi 2018). The current study is situated conceptually in this theoretical space because a goal of the research is to identify the root causes for the non-isomorphic pattern of institutionalization between federal agencies. Barley and Tolbert (1997) cite the argument of Tolbert and Zucker (1996) that

> …practices and behavioural [sic] patterns are not equally institutionalized. This variation depends, in part, on how long an institution has been in place and on how widely and deeply it is accepted by members of a collective. Institutions that have a relatively short history or that have not yet gained widespread acceptance are more vulnerable to challenges and less apt to influence action (96).

On the other hand, impulses for reform of existing practices must compete with existing institutions, if they are not mutually compatible, before they can become institutionalized. Limba et al. (2019) cite the view of Greenwood and Hinings (1996) that "…when organizations undergo change, a major source of organizational resistance derives from the embeddedness of the organization within its institutional context" and further note Jansson's (2008) observation that "…the process usually involves uprooting old institutions and installing new ones….this process is usually a tall order for a government organization because of the normative characteristics of institutions in public organization settings" (106).

While organizational change in government is normally a tall order, empirical research into public sector organizational change has found isomorphic pressures may 1) have positive effects on agency compliance but not convergence across agencies, 2) influence organizational strategies and culture but not structures or processes, and 3) unleash a level of defiance among agencies that are already highly structurated which leads, in turn, to greater heterogeneity between organizations (Ashworth, Boyne, and Delbridge 2007, 181-182). For example, Teodoro (2014) discovered normative isomorphism resulted in agencies which were led by public executives who were professionalized as engineers committing fewer violations of the Safe Drinking Water Act than those which were not (1000). Because the varying responses of U.S. federal agencies to GPRAMA may be explained as differentiated responses to isomorphic pressures generated by GPRAMA and stakeholder and oversight expectations, a sociological neo-institutional framework for interpreting evidence concerning the dynamics of organizational decision making seems well-suited to respond to the first research question.

It is therefore possible and likely fruitful to study variance in agency adoption of a reform institution (here, new rules regarding performance management practices) by analyzing variance

in the degree of institutionalization indicators and how it is affected by conditions both outside and inside federal agencies. Institutionalization "…involves the processes by which social processes, obligations, or actualities come to take on a rulelike status in social thought and action" (Meyer and Rowan 1977, 341). There is substantial terrain between what may begin as an idea for government reform and that idea eventually assuming "rulelike status" in the consciousness of federal employees and in organizational and procedural structures. Tolbert and Zucker (1996) argue "…a structure that has become institutionalized is one that has become taken for granted by members of a social group as efficacious and necessary; thus it serves as an important causal source of stable patterns of behavior" (179). They conjecture that organizational outcomes from the institutionalization of a structure likely depend on the level of its institutionalization (184). This is not meant to suggest that extensive institutionalization necessarily results in beneficial policy outcomes, merely that whatever agency outcomes do occur will depend on the extent and depth of policy reform institutionalization.

What does the organizational terrain look like through which a reform concept must pass before it evolves into normal, routine and taken-for-granted practices (i.e., an institution), and how does that terrain affect the way an organization chooses to implement the reform? From the perspective of management theory, Ansari, Fiss, and Zajac (2010) argue that, after adopting diffused practices, organizations adapt them, and such adaptation is a result of the "fit" between the practice and the organization. Drawing from Nadler and Tushman (1980), they define "fit" as "…the degree to which the characteristics of a practice are consistent with the (perceived) needs, objectives, and structure of an adopting organization" (68). Practice adaptation in organizations occurs during the diffusion process due to the effects of "…population and organization-level mechanisms" which often lead to the practice being contested (Fiss, Kennedy,

and Davis 2012, 1077-1078). Leading public management literature addresses institutionalization as one phase in a model of integrative public leadership that includes the adaptation of rules and structures, but omits leadership promotion of the reform or manager use of performance information, treating them as co-equal leadership tasks (Moynihan and Ingraham 2004, 431-432).

Organizational theorists approach the question of institutional-organizational alignment using a more comprehensive, rigorous, and well-specified method. Sarta, Durand, and Vergne (2021) critique the shorthand formulations of adaptation presented in the management literature and, based on a large meta-analysis of the organizational adaptation literature, propose a clear definition: "…intentional decision making undertaken by organizational members, leading to observable actions that aim to reduce the distance between an organization and its economic and institutional environments" (44). During the institutionalization of new practices in an organization, such actions may include employees trying to balance concurrent pressures for procedural change and program outcome consistency by adopting micro-adaptations in existing routines to accommodate the new business rules (Turner and Rindova 2012, 25). The more adaptations are needed, the most costly and difficult institutionalization is for the organization.

The institutionalization process within organizations described in the literature of neo-institutionalism offers an analogue to the intra-organizational diffusion process portrayed in management scholarship, but offers a more complete and nuanced account of the adaptation process and the dynamic effects of institutional and organizational conditions on structural changes. Tolbert and Zucker (1996) propose a refined and sequenced process model for institutionalization in which new ideas become institutionalized through three stages: habitualization, objectification, and sedimentation (180-184). According to this model,

organizations at the habitualization stage have initially responded to an innovation from outside by setting up new and independent structures with associated policies and procedures which may or may not be influenced by models in other organizations and can be considered as pre-institutionalized, in part, because consensus on the value of the new structure has not yet been reached (181). In addition, they note the number of adopting organizations may be limited, and the adopted structures may take many different forms and will tend to be impermanent (182). Reay et al. (2013) add that "habitualization draws attention to the tight relationship between activity and meaning construction, and helps to conceptualize the way that meanings develop through the enactment of behavior" (966).

According to Tolbert and Zucker (1996), organizations at the next stage, objectification, have diffused the new practice structure more widely across the organization by fostering consensus to change by assessing costs, risks, and benefits; by monitoring the success of the innovation in other organizations; and through influential theorizing by organizational champions who add legitimacy to the reform idea by clearly articulating the problem the structure can address, citing some evidence of reform effectiveness, and justifying the proposed structural solution (181-182). They characterize structures at this stage as being semi-institutionalized, somewhat less varied between organizations, and less impermanent than at the habitualization stage, but may still have a "faddish" quality to them (182). For the present study, the transition to the objectification stage is a critical one because, as they explain, it requires that organizational motivations change from imitation or simple compliance to a sense of voluntary and preliminary acceptance based on normative criteria.

Finally, Tolbert and Zucker (1996) describe organizations at the sedimentation stage as having fully institutionalized the structure, but argue the course and pace of the

institutionalization process can be affected by the influence of mobilized opposition groups or by a lack of demonstrated results from the structure. They propose that full institutionalization in the organization means structures endure over generations of members and are fully adopted by all appropriate actors (184). Sedimented bureaucratic structures involve well-mastered work routines. Such routines are beneficial because they make organizational operations more consistent and improve the efficiency of results (Turner and Rindova 2012, 25). Cohen (2007) cautions scholars to consider that routines need not be, as might be presumed from mainstream post-World War II cognitive behavioral science, "…rigid, mundane, mindless, or explicitly codified," and credits the pragmatist John Dewey and classical philosophers for contributing a more liberal perspective of routines as being more like habits, with their associations to thoughts and emotions, or learning opportunities (778-779). Limba et al. (2019) predict that achieving full and lasting institutionalization requires organization-wide learning to make innovations durable and sustainable (108). Given Cohen and Limba's perspectives, deep sedimentation of institutionalized practices is equally a product and antecedent of organizational learning and adaptation. Fernandez and Rainey (2006) argue similarly:

> Managers and employees must effectively institutionalize and embed changes. To make change enduring, members of the organization must incorporate the new policies or innovations into their daily routines. Virtually all organizational changes involve changes in the behavior of organizational members. Employees must learn and routinize these behaviors in the short term and leaders must institutionalize them over the long haul so that new patterns of behavior displace old ones (172).

Normative recommendations to institutionalize a reform policy depend on the sedimentation of routine daily practices that reflect accumulated organizational knowledge about accepted procedures that permit employees to learn by doing. Public management scholars may therefore fruitfully investigate and analyze how variance in institutionalization processes affects structures

and agency of individuals inside organizations and thereby explain non-isomorphic agency responses to reform mandates.

Baldwin, Chen, and Cole (2019) describe institutional analysis as: ". . .a suite of approaches for understanding the various ways in which formal laws and informal social and organizational norms shape policy actors' behavior," which they associate with the Institutional Analysis and Development (IAD) framework of the late Noble Laureate Elinor Ostrom (891). The IAD framework treats institutions as regulative systems with written and unwritten rules and mutual understandings (Ostrom 2011, 18). In the regulative conception of an institution, analysts ". . .assume that the major mechanism by which compliance is effected is coercion. Individuals and groups comply with rules and codes out of expediency—to garner rewards or to avoid sanctions. Behavior is viewed as legitimate to the extent that it conforms to existing rules and laws" (Scott and Davis 2007, 259). The IAD predicts that actors in an action situation will respond to incentives to achieve outcomes (Ostrom 2011, 10). In this study, federal managers who must decide how to institutionalize GPRAMA are the actors. What are their incentives, and are these always rational?

One managerial incentive often cited in the literature of sociological institutionalism is the acquisition of greater organizational legitimacy. A core and longstanding concern of institutional theory, including the classic version (e.g. Selznick 1957), is organizational legitimacy, which is defined as "…the perceived appropriateness of an organization to a social system in terms of rules, values, norms, and definitions" (Deephouse et al. 2017, 27, 32). Given its many benefits for organizations, all else being equal, the prospect of acquiring greater legitimacy for the agency or programs can presumably serve as a strong incentive for public sector managers to conform their agencies to the reform and institutionalize it. Feldman (2005)

notes that "public management is at least as much about managing the external environment as about managing the internal organization" (959). Baum and Oliver (1991, 1992), Pollock and Rindova (2003), and Singh, Tucker, and House (1986) all found legitimacy affects organizational performance and survival (as cited in Deephouse et al. 2017, 27, 35). Pfeffer and Salancik (1978) found legitimacy also supports an organization's ability to compete for resources (201). Meyer and Rowan (1977) argued under-performing organizations could enjoy protection from external sanctions if they have sufficient legitimacy (351). Knoke (1985) claimed legitimacy contributes to organizational autonomy (as cited in Deephouse et al. 2017, 36). Carpenter (2001), as a political scientist, appears to agree with Knoke and argued that "bureaucratic autonomy requires political legitimacy, or strong organizational reputations embedded in an independent power base" (14). Child (1972) cited the specific benefit of greater autonomy from legitimacy by noting legitimacy allows organizational strategists flexibility to choose strategies (as cited in Deephouse et al. 2017, 36). Public leaders certainly have an incentive to pursue greater legitimacy for their organizations.

As a significant theme of research, neo-institutional scholars have explored ways actors have sought to preserve organizational legitimacy while managing the tensions and contradictions between external pressures of institutional isomorphism and internal organizational conditions (Glynn 2017, 244). Reay et al. (2013) point to bias in neo-institutional empirical scholarship favoring research on the macro-level diffusion of ideas between organizations over micro-level tracing of institutionalization patterns within organizations (965). Ansari, Fiss, and Zajac (2010) argue the omission in the literature of micro-level analysis within organizations prevents a close examination of the evolution of diffusing management practices through a process of "…custom adaptation, domestication, and reconfiguration to make them

meaningful and suitable within specific organizational contexts" (67-68). Through the comparison of federal agencies as cases and units of analysis, this study contributes micro-level analysis that reveals how certain federal agencies adapted and reconfigured themselves to institutionalize GPRAMA.

Echoing Reay's (2013) concerns about insufficient attention to micro-level analyses of institutionalization processes, other scholars have noted most contemporary organization theory literature has overlooked studying organizations per se, preferring instead to discover context-agnostic principles for all organizations, and called for a renewal of the field's original focus of context-sensitive research on organizations (e.g. Pfeffer 2006, 459). Greenwood, Hinings, and Whetten (2014) argue that overlooking heterogeneity in structures in organizational research hinders the development of organization theory, and propose using institutional logic to study it (1207). Lounsbury (2008) suggests institutional analysts can understand the reasons for variation in organizational practices by researching the multiple institutional logics (i.e., ". . .multiple forms of institutionally-based rationality. . . .") in an organizational field (354). For example, Teodoro's (2014) findings concerning the effects of professionalism on public managers, discussed previously, suggest the presence of an institutional logic of professionalism that is separate from, and competes for managerial attention with, the institutional logic of the reform being implemented. In the chapters to follow concerning institutional and organizational conditions and how they affect agency institutionalization of GPRAMA, the effects of divergent institutional logics on political influence, agency goal clarity, agency culture, and credible leadership commitment especially will be noted where the evidence reveals possible cases in which prevailing institutional logics as a guide for managerial decision making hinders performance management institutionalization.

A related line of organizational theory, relevant to the second research question, concerns organizational identity (OI). Gioia et al. (2013) argue OI ". . .links an organization to the institutions in its environment" (154). There is a growing body of organizational research on OI as a distinct phenomenon that is embedded in, but conceptually independent from, organizational culture (Kreiner 2011). Drawing from Albert and Whetten (1985) and Whetten (2007), Kreiner (2011) offers a definition: "[OI] answers the 'who are we' question about an organization and consists of three major facets: *definitional* (institutionalized identity claims arising from central, enduring, and distinctive organizational characteristics), *ideational* (the shared beliefs of organizational members), and *phenomenological* (identity-related discourse relating to important organizational events)" (464, italics in original). Corley et al. (2006) propose that OI situates an organization within an organizational field, industry, or nation-state and permits comparisons with similarly-situated organizations (87). However, they agree with Gioia, Schultz, and Corley (2000) that OI is not corporate identity or organizational image because these are externally-facing images of the organization. Fombrun (1996) and Fombrun and Stanley (1990) note that reputation is also in the eye of external observers, whereas OI is "self-referential" (Corley et al. 2006, 86-88). Glynn (2017) argues OI refers ". . .not only to the values and ideology claimed by an organization, but also to the meaningful alignment of an organization's defining attributes or character ('who we are') with its core practices and behaviors ('what we do')" (244).

The parallels of OI with the administrative leadership concepts of Selznick (1957) are immediately apparent; as Glynn observes, however, Selznick conceived of OI as an outcome of prior institutional processes (Glynn 2017, 243). The perspective in this study treats OI as both an antecedent and a condition that affects the progression of institutionalization in organizations. The literature on the formation of OI, as an antecedent to institutionalization, is less fulsome than

that of OI change, but it seems to be adequate to theoretically ground this portion of the study. For example, empirical research into new and successful European universities found that the initial formation of their OI may have been influenced by a desire by leaders to improve their legitimacy by modeling themselves after existing universities in the same organizational field (Czarniawska and Wolff 1998 as cited by Gioia et al. 2013, 155). Case study research into the establishment of a new science and technology school revealed that initial claims to OI were directly influenced by the vision and goals of its founders (Gioia et al. 2013, 158). The US Congress is the founder of all U.S. Government agencies, and establishes the vision, goals, budget, and programs the agencies will run. These are all well-documented, published in laws, and reaffirmed and monitored through oversight mechanisms and recurring agency testimony. It is therefore extremely probable that the core of the OI of federal agencies, at least Kreiner's definitional element, derives from its legislative mandate. Legal mandates theoretically imprint themselves on OI; however, complex mandates may make an agency's identity ambiguous (Gilad 2015, 605).

OI is not necessarily permanent. Indeed, the OI literature has shown that it is socially constructed by internal and external stakeholders, therefore mutable (Hatch and Shulz 2002 as cited by H. Doering et al. 2021, 377). OI scholarship has found that OI can change, even over short periods (Gioia et al. 2013, 126). But organizations are widely known to resist change, especially when it may conflict with OI. Kodeih and Greenwood (2014) argue that ". . .if an organization's identity is inconsistent with institutional prescriptions, the organization will resist and/or revise those demands" (10). In such cases, OI may mediate the influence of external change pressures on the allocation of managerial attention to organizational tasks. In the case of public organizations, those with a clear OI were found in comparative case studies to be more

likely to have clearly-prioritized and deeply-institutionalized goals[6] that led them to initially resist external tasks inconsistent with its OI through the use of superficial or symbolic actions to respond to the tasks, but then assimilated the new tasks into their OI over time in order to maintain their legitimacy (Gilad 2015, 593, 605). According to Gilad (2015), organizations with an ambiguous OI will tend to have an unclear prioritization of tasks and goals and, all else being equal, be more susceptible to external pressures to focus attention on the new tasks, resulting in these agencies giving them higher priority and being early and substantive adopters of the new tasks (605).

Agency responses to reform may therefore be motivated in part by a desire to protect the organization's OI from existential change and its legitimacy from reputational risks. Notwithstanding the conceptual and definitional separation of organizational identity and reputation described above, feedback concerning organizational reputation can validate or threaten an organization's image of itself. Research evidence points to an acceleration in organizational change when managers perceive a discrepancy between the organization's reputation and its OI when its OI is more favorable than feedback about its reputation (e.g., Martins 2005). Actively participating in a performance management regime may be a solution for an agency to reconcile these divergent images. According to H. Doering et al. (2021), OI research in the public sector has shown that "voluntary and proactive engagement with new forms of performance assessment is…also an opportunity for active reputation management" (377). That opportunity may appear more enticing to organizations with an ambiguous OI than to those with a clear OI, because the latter may perceive active participation in the

[6] Gilad's (2015) conceptual model integrates goal clarity as an element of OI (596). This study treats OI and goal clarity as separate organizational conditions based on their treatment in the public management literature.

implementation of the reform offers them less marginal potential benefit and greater relative risks to their reputation and legitimacy than the former do, because agencies with ambiguous OIs rely much more on their reputation to protect their legitimacy. As Carpenter and Krause (2012) cogently propose:

> . . .differentiation [in functional composition of agency workforces] means that agencies are harder to characterize, which means (counterintuitively but powerfully) that agency reputation is all the more important because organizational images offer forceful simplifications of more complicated agency realities, *and they are often relied on more heavily when the agency is more complicated*" (28, italics added).

Such is the power of deeply-institutionalized practices: in time, they come to be perceived as inherent to an agency's identity. Reform ideas that could change or discard those practices in favor of new ones can threaten OI, so defensive responses from agencies with clear OIs that do not align well to reform tasks should not be unexpected. Those with complex or murky OIs based on the complexity of their legal mandates and functional profiles should put up less resistance to top-down reform mandates, like GPRAMA, because their OIs depend less on a set of deeply-institutionalized tasks at risk of being changed. This argument, of course, excludes the significant influence of the other external and organizational conditions that affect agency institutionalization of reform practices described in detail in the next section. OI is therefore but one of several conditions that may affect GPRAMA institutionalization.

In light of this theory and evidence, all else equal, OI may affect the rate and extent of institutionalization of government reform in federal agencies, and one may predict that agencies with a clear OI and a well-institutionalized task prioritization that conflicts with the tasks of the government reform may fall behind other agencies in the institutionalization of practices associated with the implementation of the government reform. Tasks to institutionalize performance management practices require management attention. OI should thus condition the

allocation of federal agency management attention on GPRAMA institutionalization. Over extended time periods (perhaps decades as in the case of GPRA and GPRAMA, as chapter 3 will describe), as a result of isomorphic pressures including the struggle for legitimacy and positive reputation, even slow agency adopters will ultimately find a way to adapt and institutionalize many or most structures related to the reform, as long as institutional conditions are aligned in a way that continues to sustain steady pressure on agencies to change. Based on theoretical predictions in the literature, the less ambiguous an agency's OI, and the less commensurate its institutionalized routines with the reform are, the longer its adaptation to the reform will take and the more costly it will be, all else being equal.

Because OI offers a theoretical criterion for differentiation and comparison of similarly-situated organizations in an organizational field, it should permit the comparison of bureaucracies by their type per Wilson's (1989) taxonomy of bureaucracies. Federal agencies possess unique organizational identities in terms of their legislated functions (what they do) and, over time, have developed a sense of institutionalized self-awareness that includes cultural characteristics such as norms and values (who they are). Assuming bureaucratic type is a valid proxy for OI in public sector organizations, the description and analysis of the effects of bureaucratic types on reform institutionalization in chapter 5 offers empirical evidence that demonstrates the force of OI as a mediating condition on this study's primary indicators of GPRAMA institutionalization.

Moving forward, this study employs the "cultural" perspective of neo-institutionalism throughout, and uses historical institutionalism to explore the rich, complex, and fascinating history of the modern performance management movement in the U.S. federal government, and a recent expansion of sociological institutionalism that focuses on organizational conditions to

analyze agency and federal manager responses to GPRAMA requirements in chapter 4. In chapter 5, it also analyzes patterns of variance in the institutionalization of GPRAMA practices based on agency identity and bureaucratic type through the lens of "old institutionalism" which underlies James W. Wilson's taxonomy of bureaucracies.

Public Management Scholarship on Institutional and Organizational Conditions

Chapter 1 framed the main public management dilemma of this study: why is there variance between agencies in their institutionalization of a legislatively-mandated government reform? What conditions in the governance environment and inside departments and agencies contribute to this variance, which creates significant legislative oversight and democratic accountability concerns? As described in the previous section, sociological neo-institutional theories conceptualize institutionalization as the degree of organizational embeddedness and normative internalization of structures and routinized practices. They propose that coercive, mimetic, and normative mechanisms of institutional isomorphism tend to homogenize patterns of organizational institutionalization; and suggest that variance in agency institutionalization can be attributed to the mediating effects of organizational field and organizational conditions, including OI, on the habitualization, objectification, and sedimentation of new practices in agencies. Similarly-situated government agencies in an organizational field, such as all CFO Act agencies as in this study, respond to the three mechanisms of institutional isomorphism, but not equally or identically so. Agency responses which impede institutional isomorphism of policy reform implementation approaches offer evidence that public managers make choices about how and when to adapt their organization to the reform. Precepts of neo-institutional theory suggest these choices are shaped by a set of incentives common to leaders and managers in all public organizations. For public agencies in which financial gain is generally not a consideration, the

preceding section proposed several: organizational survival, legitimacy, autonomy, and the protection of OI.

Another potential factor for some agency leaders is a personal or professional agenda based on policy entrepreneurialism. In the public administration literature, Moynihan, Pandey, and Wright (2012) summarize the extensive literature on transformational leadership to observe that advocating for the achievement of organizational outcomes is characteristic of transformational leaders (146-147). The prospect of improved organizational performance may therefore create incentives for managers when agency leadership is transformational. I argue that incentive structures, i.e., combinations of incentives that managers experience, depend on the relative salience of each of these incentives to the manager based on the relative presence or absence of agency-specific conditions. Furthermore, I draw on research that these incentive structures interact with managers' intrinsic public service motivation (e.g. Georgellis and Tabvuma 2010; Kroll and Vogel 2014), career ambitions (e.g. Liang 2014), risk tolerance (e.g. Hood 2011), role identity (e.g. Pfiffner 2019), general attitude towards reform (e.g. Kroll 2014), and a host of psychological biases (e.g. Battaglio et al. 2019; Belle, Cantarelli, and Belardinelli 2018) to drive strategic decisions about how and when to adapt organizational structures and practices to the requirements of the reform.

Modeling all of the complex and dynamic psychological interactions between organizational conditions, incentive structures, and managers' psychological characteristics at the micro-level is beyond the scope of this study. Pollitt (2013) observes that public administration scholars rarely invoke the literature of psychology to study public management decision making (348). Some elements of the present study, however, intersect with a recent and

thriving literature on Behavioral Public Administration (BPA)[7] that leverages the discipline of psychology and behavioral theories of administration (e.g., Grimmelikhuijsen et al. 2017). Empirical findings of manager responses to survey questions and interviews in the present study regarding their intrinsic motivation and attitudes towards performance management offer useful qualitative evidence and insights for larger BPA debates and ensure a behavioral perspective informs the present research.

In this study, federal agency manager decisions to embed and sustain new structures and practices in daily routines that enable their organization to satisfy GPRAMA requirements constitute observable indications of past performance management reform institutionalization decisions at the organizational level. What contextual and organizational conditions affect these decisions, and in which configurations? With so many conceivable influences weighing on public managers' attention, a sufficiently parsimonious yet appropriately specified inferential model, informed by a review of the public management literature, needs to be developed and tested in order to generate meso-level findings useful for informing policymaker and senior public manager strategies to influence their organizations and organizational environments in a way that fosters deeper institutionalization in public agencies. This study accepts that challenge.

As a basis for a new inferential model to analyze the relationships between system and organizational conditions and agency institutionalization outcomes, this study adopts the Multiple Governance Framework (MGF) (Hill and Hupe 2006), an extension of the IAD framework tailored specifically to the public policy process. Table 2.1 shows this framework.

Table 2.1. Multiple Governance Framework.

	Action Level

[7] For a current and well-developed overview of BPA scholarship, see Bhanot and Linos (2020).

Action Situation	Constitutive Governance	Directive Governance	Operational Governance
System	Institutional Design	General Rule Setting	Managing Trajectories
Organization	Designing Contextual Relations	Context Maintenance	Managing Relations
Individual	Developing Professional Norms	Situation-Bound Rule Application	Managing Contacts

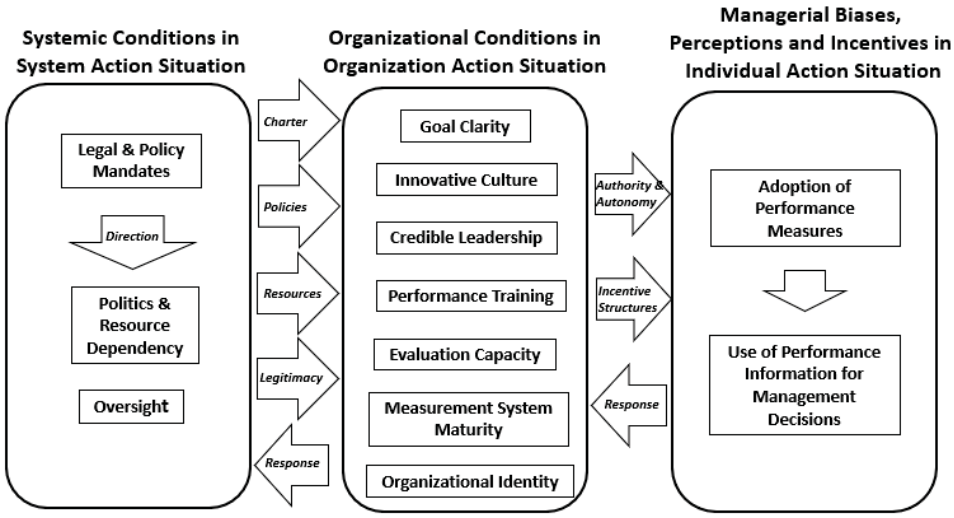
Source: Hill and Hupe (2006), 563.

In this study, I am concerned principally with variance in agency responses to a major public policy reform as a consideration for reform policy implementation, not larger questions of constitutional design, coequal interagency relationships, or micro-level (operational) management concerns. Therefore, the study eschews questions of constitutive and operational governance and focuses on the directive governance action level. Here, the Congress, federal oversight entities such as the GAO, and the federal interagency community[8] led by OMB constitute the system action situation. They set and enforce the rules of the reform and monitor its institutionalization in federal agencies. In theory, at least, they have a substantial degree of coercive power through the force of law, policy, and resource dependency. Individual federal agencies constitute contexts in the organization action situation. Each agency is its own context and, as all organizations do, maintains that context through the maintenance of its OI; its strategic plans including goals and objectives; its internal rules codified in policies, procedures, and standards; effective management of its resources (including its human capital and technologies); organizational narratives; and other structures. Finally, public manager[9] decisions

[8] Through the lens of institutional theory, the collection of all CFO Act agencies in the system action situation constitutes an organizational field subject to institutional forces from Congress.
[9] As conceptualized here, "public manager" in nearly all cases does not refer to an individual manager *per se,* but to a collective and collaborative network of managers within a federal agency which, together, make sense of the reform requirements and make decisions concerning the institutionalization of performance management practices in their organization. In this respect, "individual action" is still very much collective action.

that affect agency institutionalization of practices supporting GPRAMA requirements, including the selection of performance measures and decisions on how performance information will be used for management decisions, are framed within the "individual action" situation. Overlaying the contextual conditions that mediate the performance management policy institutionalization process with the MGF, this study proposes a new inferential model for analysis that integrates these conditions across the system, organizational, and individuals action situations in the MGF at the directive governance action level, as shown in figure 2.1. below.

Figure 2.1. Inferential Model for Performance Management Policy Implementation



Source: Author's own illustration.

In this framework, the system action situation is an "institutional environment" and conceived to be a complex regulative system as defined by Scott and Davis (2007, 258-259) that requires organizational compliance with its legal and policy mandates. In the system action situation, the implementation of legal and policy mandates is mediated by the systemic conditions of political pressures and agency resource dependency, which the literature generally treats as a combined condition because, as Wildavsky and Caiden (2004) note, resource allocation and budgeting decisions are usually driven by political priorities, and by the form and

intensity of oversight mechanisms. Mechanisms of formal oversight, such as OMB, inspectors general, and the Government Accountability Office (GAO), are likewise situated in the system action situation.

In the inferential model, organizational conditions are set within the organization action situation. Organizations exist in a principal-agent relationship with Congress (the principal) and external oversight authorities, and depend on them for their legal charter, policy direction, resources, and legitimacy. Citing the literatures on policy diffusion and public sector innovation, Andersen and Jakobsen (2018) propose top-down political pressures and the power of institutional isomorphism theoretically explain large-scale innovation adoption by agencies, but point to organizational conditions as more probable causes for variation in agency adoption of innovation (216). In this study, both are assumed to exist at the system action level. This study accepts the logical implication of this prediction that both system action and organizational conditions affect agency institutionalization of performance management reform, but without ruling out the added possibility that system action situation conditions may also directly affect variance in agency-level institutionalization.

Through changes to agency charters, policy decisions, withholding or granting resources, or questioning agency or program legitimacy, the principal can influence the level of coercive, normative, or mimetic isomorphism agencies experience. In the context of the U.S. Government, the Congress exercises significant legal power over federal agencies. For example, communicating threats to agency or program legitimacy can encourage failing agencies to emulate prominent exemplars. Senior agency leaders and managers, or management teams, are set within the individual action situation and are responsible for choosing whether and when to adopt performance measures and use performance information for management decisions. Their

individual authority and autonomy to make decisions and allocate organizational resources, as delegated by their organization, define the outer bounds of their decision-making opportunities, and their incentive structures permit managers to select courses of action within the scope of their authorities which fit best with their incentives. Furthermore, managers' attitudes towards change and their intrinsic public service motivation mediate the effects of the incentive structures. The central claim of this study is that either individually or in some causal configuration, the conditions in the system, organization, and individual action situations in this model mediate the institutionalization of reform policy in public organizations.

The following literature review examines each condition of the inferential model in detail. It proceeds with a critical evaluation and synthesis of the theoretical perspectives and empirical research concerning the theorized influence of each condition on performance management institutionalization. For each condition, it also proposes testable hypotheses related to the research questions, and explains how the study models and operationalizes the condition. Chapters 4 and 5 describe the research design, data sources, and quantitative and qualitative analytic strategies employing these conceptual operationalizations.

Conditions in the System Action Situation

Politics and Resource Dependency

In this study, external political pressures are defined as actions based on coercive or normative authority, excluding legal and policy mandates and official oversight activities, from the Congress, other Executive Branch entities, or public interest groups that are designed to encourage agency leaders to give priority in decision making to an external political agenda over the purely functional requirements of the agency and its programs. Internal politics is excluded from this definition. As Yang and Hsieh (2007) note, ". . .performance measurement has to be

understood as embedded in bureaucratic politics" (871). Ho (2007) expands on this view by noting the record of implementation of federal performance management reforms has demonstrated that the governance environment and power dynamics in the policy development process often create significant challenges (370). The influence of politics, often coupled with budgeting decisions which are inherently expressions of elected leaders' political preferences (Wildavsky and Caiden 2004, 20-21), may induce agencies, under a favorable alignment of conditions, to institutionalize government reform legislation. Empirical research supports this claim. For example, based on the results of two survey experiments with school principals in the U.S. and Denmark, Andersen and Jakobsen (2018) found political pressure had a significant and consistent effect on the willingness of public managers to devote resources to mandated organizational innovation (234).

Political influence is exercised in the framework of a principal-agent relationship between the legislature, as the principal, and federal agencies, as agents. Public managers exercise their responsibility for making agential decisions and, as agents, they enjoy discretion arising from under-specification of factual and value premises for their decisions by a principal (Simon 1997, 308). In the context of a public bureaucracy, ambiguity in policy purposes and means creates discretion for managers (Moore 1995, 63). Many conditions mediate that relationship, and therefore also the impact of political influence on agency institutionalization decisions. This study sides with the critics of the concept of the politics-administration dichotomy, first propounded by W. Wilson (1887), and argues that how much power the legislature, as the keeper of the public purse and source of all laws, holds to drive the institutionalization of reform policies in American federal bureaucracies is mediated by the presence or absence of organizational conditions in agencies, as well as federal manager

discretion. Politically-sensitive or highly-impactful functions of an agency can increase the degree of congressional control regardless of other agency relationships with political leaders (De Jong 2016, 58). On the other hand, political pressure may also create perverse incentives that induce public managers to use performance management to create false impressions of agency performance, engage in gaming behaviors, or misalign their agency's performance goals with the government's policy goals or the agency's public context (Cuganesan, Guthrie, and Vranic 2014, 284). There is never an *a priori* guarantee of alignment in views on performance management between political principals and agency senior leaders. Moynihan (2005) observes ". . .as elected officials impose a reform on agency officials, each group acts in a way consistent with its motivations, norms, and interests" (221).

Trust between politicians and agency leaders, especially when both participate in a policy community, positively mediates the institutionalization of policy (Miller and Demir 2007, 141). Actual agency performance has been shown to be positively associated with the degree of alignment of views between political leaders and agency senior managers on performance management when agency leaders trust politicians (Walker, Jung, and Boyne 2013, 841). The converse may also be true: agency leader distrust of politicians or political appointees may hinder reform institutionalization. Wood and Lewis (2017) note a recurring finding in the literature that ". . . in the United States the increased 'presidentialism' of the administrative state is generally viewed in a negative light because of its effects on agency performance and 'neutral competence'" (582). Politicians who are not sincerely interested in good government may support performance-related agendas not to improve agency performance per se, but to pursue potential political payoffs with relatively little political risk (i.e., changing internal agency management plans and processes is not likely to offend, or even be noticed by, many

constituents). As Kingdon (2003) argues, these payoffs could be gained through favorable publicity for their advocacy for a policy with little controversy and high salience, passing redistributive laws that benefit constituents, or burnishing their Beltway reputations as an agenda-setter (38-39). As chapter 3 shows, performance management reform, framed as a "good government" initiative, is usually politically noncontroversial and, as in the case of GPRA and GPRAMA, generally enjoys bipartisan support, although for different reasons.

Performance reforms operate on the theory that performance information should influence distributive decisions between programs in the budgetary process. A legislator who can influence the budget using performance data in favor of preferred programs, especially those that benefit his or her constituency, may reap electoral payoffs. Similarly, one who wishes to build a reputation as a moderate who works collaboratively with the other party may find participation in performance-related committees to be a good professional opportunity to grow such a reputation.[10] As long as the reform serves such instrumental purposes for legislators, political support will likely be sustained. Once it does not, political support may quickly erode unless the legislator has a sincere personal commitment to high-performance government. Deciding to undertake a major public reform can entail significant risks for a political leader:

> Reform outcomes are seldom certain and poor reform outcomes are likely to attract considerable blame to the responsible political decision makers. Adding to this, reforms are often unpopular among important stakeholders–particularly among the reformed organizations and their clients–who may attempt to fight back, resulting in immediate risks of attracting blame. Undertaking reform is therefore a risky endeavor that requires a balancing of the blame risks involved in the pursuit or nonpursuit of reform (Nielsen and Baekgaard 2015, 552)

[10] This bipartisan/technocratic reputational positioning strategy worked well for rising stars in Congress like Albert Gore in the 1980s and 1990s (Heilemann 1995). In today's tribalized political environment, it would likely leave the legislator isolated and vulnerable to attacks about a lack of party loyalty.

Given the prevalence of negativity bias amongst legislators, typically reinforced by similar biases in the media, good public agency performance is rarely noticed and reported agency performance failures cause politicians to shift blame to bureaucracies, thereby creating incentives for risk aversion and disincentives for risk taking among public managers (Hood 2011, 11). However, experimental evidence from local governments in the European Union supports the claims of Nielsen and Baekgaard about blame avoidance strategies used by politicians; it suggests information demonstrating low or average performance is positively correlated with preferences for increased spending from politicians from low-performing municipalities, but information about high performance has no statistically-significant effects on spending preferences (George et al. 2017, 1262-1263).

This study integrates agency resource dependency into the system action level political control condition because, as conceptualized by resource dependency theorists, control of resources confers a measure of power on the controller over the dependent recipient (Pfeffer and Salancik 2003, 258). Braun and Gilardi (2006) frame the use of contingent resourcing as a tool of coercive political influence to impose a change in policy (310). A recent bibliometric analysis of the resource dependency literature over the past 45 years documented extensive use of resource dependency theory by scholars of strategic management, and revealed the two most-often used keywords in management journal articles related to resource dependency were "performance" in 29.7 percent of articles and "management" in 17.2 percent (Ozturk 2020, n.p.). Though it is separate from organization theory, the nexus of resource dependence theory with the current study is therefore intimate. Malatesta and Smith (2014) argue public managers attempt to affect the power dynamics in agency relationships in a way that maximizes agency autonomy by minimizing external dependence, and try to "leverage influence (by controlling resources that

others need)" (16). They propose three main categories of strategies public managers may consider, including mergers, alliances, and co-optation (17-21). For federal agency responses to legislated performance management reforms, only the latter strategy seems potentially applicable. Arguably, it occurs in the form of active engagement by a resource-dependent agency in interagency and legislative venues with OMB and Congress to try to shape their expectations that the agency is cooperating, and to seek their assistance when it has difficulties. It is probably hard for political leaders to threaten to withhold resources from an agency they perceive to be working hard to correct its shortfalls.

The literature has documented cases where resource considerations have influenced the extent of institutionalization of government reforms under different conditions. Because neo-institutional theory suggests institutionalization of reform requires organizational learning to advance, scholars have explored the relationship between resource slack and organizational learning. Moynihan and Landuyt (2009) found a significant positive association (1101), which may mean that resource abundance positions agencies which would otherwise desire to participate in the reform to be more responsive than similarly-intentioned agencies with relatively less slack. But that assumes managers will choose to apply the agency's abundant resources towards employee learning, which is not a certainty. On the other hand, Hong (2020) found localities in Korea that were less well-endowed and more resource-dependent on the central government were more likely to institutionalize performance management reforms to cut red tape because the central government, through its control of resources, was able to exert relatively greater power over them than over less-dependent localities (942). Recent experimental research into the use of behavioral interventions targeted to public managers demonstrated that even small changes in agency budgets conditioned on their choices to

implement a reform can affect managers' responses to adopt performance measures and use performance information, possibly because public managers interpret small budgetary incentives as a signal of the importance of the reform to political principals (Andersen and Hvidman 2021, 236).

Political influences have been shown to affect specific aspects of performance management institutionalization in agencies and at the OMB level. As described in chapter 3, Gilmour and Lewis (2006) found political preferences of elected officials biased the selection of programs subjected to more intense performance scrutiny during budgetary reviews (750). Agency politicization (Wood and Lewis 2017, 581); political salience (Rainey and Jung 2010, 41); legislative ambivalence (Benjamin and Posner 2018), neglect (Cook 2014, 225), drift (Schillemans and Busuioc 2015, 201), "forum paralysis" (Maggetti and Papadopoulos 2018, 175), and interagency relationships (Andersen and Jakobsen 2018, 214) have all been hypothesized to influence the institutionalization of performance management and other accountability reforms in different ways. Christensen et al. (2018) showed that priorities for performance indicators by political principals are conditioned by their public versus private governance preferences and may change based on their interpretation of performance results; specifically, if public managers perceive that the governance preferences of principals trump their sustained commitment to policy goals, then their incentives to measure performance could be weakened (206-207).

When legislative leaders demonstrate an ambivalent attitude towards performance information reported by agencies, federal managers call into question the credible commitment of some legislative leaders to the reform. There are, nevertheless, plausible theoretical reasons why legislators might appear to be disinterested in agency performance reports. In some cases,

the information may simply be too sensitive politically for political leaders to use in policy making and budgeting, and so it may be ignored or receive special handling. For example, Light (1997) cites the formation of a special commission to consider the closure of military bases as an example of a protective Congressional reaction to performance information on some failing military programs (233), presumably rife with electoral consequences. Ideology biases politicians' evaluation of performance information as well. When performance information contradicts legislators' governance preferences, legislators will ". . .choose to 'shop around' for the dimensions of performance that can best be used to justify their prior governance preferences regarding public versus private service provision" (Christensen et al. 2018, 209). Carpenter (2001), as cited by Maggetti and Papadopoulos (2018), referring to general legislative oversight of regulative agencies but not performance management specifically, claims legislators may have a political incentive to refrain from intervening into agency affairs because doing so could foreclose future opportunities to shift blame to them, which could potentially damage their own official reputation if they are seen as complicit in the regulatory failure (180). If Carpenter is correct, the oversight dimension of performance management reform represents a potential golden opportunity for politicians to shift blame to agencies and avoid political accountability for government failures. They can criticize under-performing agencies, and their leaders, based on unfavorable agency performance reporting without exposing policy design, or their own performance, to closer scrutiny.

Public managers, of course, have their own interests and reputation to protect. Research on policy diffusion[11] suggests public managers perceive a range of external forces, including

[11] Strictly speaking, this study focuses on policy implementation within one government, not policy diffusion between governments. The range of incentives confronting public managers to

political pressures, that shape the incentives they respond to when implementing policies (e.g. Braun and Gilardi 2006; Graham, Shipan, and Volden 2013). Other external pressures include mimetic pressures to conform to effective performance management models, match performance gains at other agencies, or emulate performance management standards of professional organizations; likewise, a perceived need for greater democratic legitimacy, financial resources, and juridical-based authority may incentivize participation in a reform (Andersen and Jakobsen 2018, 219). Empirical research in the Department of Defense demonstrated that incentives to implement NPR reforms during the Clinton Administration may also have been associated with federal managers' evaluation of whether or not the reform represented an opportunity for gaining internal political leverage by using the reform to justify internal policy changes they preferred, a kind of "weaponizing" of the reform mandate (Durant 2008, 283, 290). So, there is a case for the claim that external political pressures and internal strategic agendas can generate positive incentives for federal managers to institutionalize reform.

On the other hand, the politics of performance management reforms can generate strong disincentives for federal managers by potentially exposing their agencies or themselves to professional and reputational risks (as previously mentioned, this could be an incentive for legislators). Cuganesan, Guthrie, and Vranic (2014) argue one risk of performance measurement practices is the possibility of harm to public employee welfare, which increases when performance measurement systems are poorly designed, implementation is top-down, and "…reforms [are] biased by economic logic, displacing other performance dimensions" (282, 285). Nielsen and Moynihan (2017) found interest groups can influence elected leaders to

adopt and institutionalize policies is, nevertheless, very similar in each field and so it is valid to mention them here.

interpret performance information in a way that suits their political agendas, prompting "motivated skepticism" by elected leaders to discount performance information that does not align to those agendas (270-271). Politicians may then seek to deflect public blame for agency or program performance shortfalls on to individual public managers whom they try to hold accountable for them, even if achieving the policy outcomes is well beyond their legal authority and responsibility (Nielsen and Moynihan 2017, 277). But attempting to exercise strong political control on public managers may create adverse consequences for performance management regimes. Pfiffner (2019) found a high level of political control is a significant predictor of public manager role conflicts and limited use of performance information (423). This finding appears to illustrate Kettl's claim, previously cited, that public managers often see performance management as a "gotcha game" because political leaders can and do claim credit when programs perform well, but will cast blame on agencies and bureaucrats when they do not.

In addition to contradictory incentive effects of political influences, the overall effects of politics on the institutionalization of performance management reforms may well be overestimated in the literature. Some research has found that internal organizational conditions, such as bureaucratic values, exert a stronger influence on agency outcomes than political pressures do (Meier and O'Toole 2006, 188). Furthermore, both political leaders and federal managers tend to expect few payoffs for good performance results and try to deflect blame for bad performance, thereby creating incentives for both groups to make accountability mechanisms as complex and opaque as possible. Hood (2011) argues that "crooked and ambiguous accountability trails may not serve democracy or good governance. But they can protect the political and administrative class from blame after failure. In a society whose politics and government exhibited strong negativity bias, isn't that precisely how we would expect politicians

and bureaucrats at every level to behave?" (12). A clear mandate for performance reform may not always get public managers' entrepreneurial juices flowing without some protection from the potential professional consequences for "poor" performance.

In sum, the literature points to political influences and resource dependency having some association with the adoption of performance measures and the use of performance information, although which measures are adopted and how performance information is used and reported may be hindered by conflicting political priorities or blame avoidance. Agencies with resource slack may be well-positioned to learn new performance management methods. They may be less motivated, however, to institutionalize reform mandates than agencies in greater need of resources more beholden to the central government or, if a federal agency, to the legislative principal. The literature also shows the effects of political influences are mediated by a wide range of organizational conditions that vary between agencies. Because the range of possible valid indicators for measuring political influence and resource dependency is vast, this study specifies models including nine relevant indicators employed in prominent/published scholarship as the basis for developing a factor variable measuring the condition:

1) Agency leaders perceive policymakers have substantial influence on their agency (Moynihan and Hawes 2012, S101).
2) Agency leaders rely on partnerships and networking to manage their agencies (Moynihan and Hawes 2012, S101).
3) Agency leaders must respond to disparate external stakeholder groups which are in conflict with each other over policy priorities (Moynihan and Hawes 2012, S101).
4) Political leaders expect that agency leaders will emulate the performance management systems of other agencies (Andersen and Jakobsen 2018).

5) Political leaders expect that agency leaders will improve their agency's performance to match that of other agencies (Andersen and Jakobsen 2018).
6) Political leaders expect that agencies will conform to published standards of professional associations for performance management (Andersen and Jakobsen 2018).
7) Agency leaders believe reporting performance is required to justify their budget requests (De Jong 2016, 32; Shand 2007, 95).
8) Political leaders expect agency leaders to use performance information to hold individual public managers responsible for achieving policy or program outcomes, regardless of their responsibility for them (Nielsen and Moynihan 2017).
9) Political leaders are influenced by interest groups to advocate for certain measures of performance and to interpret performance information in a way that suits their political agenda (Nielsen and Moynihan 2017).

This study models the construct of political influence and resource dependency as a condition of the system action situation. It operationalizes this condition as a fuzzy set that equally weights these nine indicators of the presence of political pressure and resource dependency. Use of a fuzzy set facilitates an analysis of whether an agency is outside, partially in, or fully in the set of agencies which experience strong political influence and resource dependency. Additional detail on the data source and measurement strategy is provided in chapter 4. To test the primary claims in the literature on the influence of political influence and

resource dependency pressures on performance management reform implementation, two preliminary hypotheses are proposed for exploratory research.[12]

H_{1a} Political and resource pressures are associated with the adoption of performance measures in agencies.

H_{1b} Political and resource pressures are associated with the use of performance information for management decisions in agencies.

Oversight

Oversight is defined here as legislated legal, policy, regulatory, and financial controls over an agency and its programs by external entities which are designed to ensure accountability to the public. It differs from political influence as a system action level condition because it is not motivated by any party or political agendas, merely public accountability and the neutral enforcement of law or policy. This is not to suggest that oversight evaluations are of no political interest to stakeholders of reform. The history of the modern performance management movement in the U.S. Government in chapter 3 clearly shows this is not the case. Agencies, however, should have less reason to be suspicious of political motives behind oversight activities than they are of overt or hidden political influences in legislative initiatives. They should have every reason, however, to be concerned with the potential repercussions from adverse evaluations of agency institutionalization of reform in oversight reports, if they anticipate political leaders will read them and act on them.

[12] Due to conflicting and roughly balanced theoretical propositions and empirical findings in the literature presented in this section, these hypotheses are preliminary because they make no *a priori* assumptions about whether the relationship is positive or negative. It could be either when combined with the presence or absence of other conditions in a causal configuration. The net effect of the condition on manager incentives could, of course, also be neutral.

In the principal-agent relationship between federal agencies and Congress, the presence of strong oversight of federal agencies by the GAO, as the legally-assigned oversight entity under GPRAMA, as well as by OMB and Congressional committees, should theoretically make a great deal of difference in agency institutionalization of GPRAMA. Oversight requires a high level of transparency and accountability to Congress and the American public. As portrayed in sociological neo-institutional theory, agency responsiveness to oversight authorities should strengthen agency legitimacy by showing legislators and the public that the agency is transparent in its operations and is trying to conform to reform requirements. Yet, chapter 1 presented clear evidence of widespread non-implementation of GAO recommendations on GPRAMA implementation by federal agencies, suggesting the impact of oversight on GPRAMA institutionalization may be muted. Why do some agencies respond so timidly, or not at all, to GAO recommendations? One possibility, suggested by Schillemans and Busuioc (2015) as cited by Maggetti and Papadopoulos (2018), is that oversight entities "…do not on their own constitute 'a real principal' which has delegated authority and has a stake in delegated regulatory activities. Instead, they receive delegated monitoring powers from the original principal. Therefore, these forums cannot be expected to have a direct interest or incentive in controlling the agency" (175). This leaves legislators as the only truly interested stakeholder group, and as shown in the previous section, legislators are often ambivalent about agency implementation of reform policies, despite public statements to the contrary. Therefore, agencies, expecting little political blowback from oversight reports, may not perceive any real threat to their legitimacy, reputation, or autonomy by refusing to fully implement oversight guidance.

There are practical limits to effective oversight as well. Programs which are outsourced to private entities, or are jointly administered with state or local governments, create significant

oversight challenges as non-federal performance goals may "attenuate" federal goals and "fragment" federal programs (Dubnick and Frederickson 2009, i156). In addition, outsourced and contracted federal programs complicate agencies' efforts to collect, analyze, and report performance information (Frederickson and Frederickson 2006, 176). As a result, the influence of oversight on GPRAMA institutionalization should be more salient to federal managers in agencies which directly administer their programs and execute the associated resources without assistance from third parties or networks; conversely, oversight should be less salient a condition for managers in agencies which contract many programs out, because programs in these agencies are inherently difficult to oversee, which affords public managers an avenue for deflecting blame to the decentralized governance arrangements. It should not be ruled out that, in some extreme cases, some agency senior leaders may take advantage of these limits to strategically deflect or downplay blame for inadequate institutionalization of reform policies.

Nevertheless, the consistent perspective of sociological neo-institutionalists about the importance of sustaining the legitimacy of public organizations, discussed previously, suggests public managers will try to avoid putting their organization's legitimacy at any real risk. Oversight can be considered a regulative pressure, and such pressures, if structured in ways that emphasize agency compliance, may hinder full institutionalization (Maggetti and Papadopoulos 2018, 232-233). This is because, as claimed by a 1999 National Academy for Public Administration study, ". . . accountability focused on compliance with administrative procedures can interfere with accountability for achieving results" (National Academy of Public Administration 1999, 5). Empirical evidence supports this view. Moynihan and Lavertu (2012) found that conflicts between oversight authorities and agencies can result in federal managers instituting performance management work routines that formally adopt performance measures to

appease external stakeholders, but not to use performance information for management decisions as the legislature intended (598-601), thereby satisfying minimal requirements for compliance. Performance management oversight also appears to be limited to spending programs so far. Use of accountability mechanisms also appears to vary depending on whether the program is a spending program or a tax expenditure. Based on GAO data, Benjamin and Posner (2018) found that, while tax expenditures aim to achieve policy outcomes just as spending programs do, policy principals do not hold agencies to the same level of accountability for the results-oriented performance of tax expenditures under their management as they do for their administration of spending programs (580). They also observed that neither OMB nor the Treasury Department have exercised oversight over performance measurement concerning tax expenditures as required by GPRAMA (570).

New Public Management (NPM) philosophy proposes that agencies which are accountable for their programs and reliably report their performance to their principals should enjoy greater trust and autonomy. Agencies which are less accountable and unable to track their performance will enjoy less autonomy and more oversight. Federal managers may not desire the additional engagement from oversight authorities, as it could expose them to politicization and goal conflicts. In general, public managers have less managerial autonomy than private-sector managers do (Swiss 2005 and Chen 2007 as cited by Hvidman and Andersen (2014), 38-39). Typically, however, they seek to defend and promote the interests of their organizations (Selznick 1957, 58) and to increase the relative institutional autonomy of their agencies to secure their ". . .distinctive values, competence, and role" (Terry 2003, 50). We see this in practice, for example, in Europe, where greater legal autonomy for public organizations and lower levels of

politicization were found to predict the institutionalization of NPM-style reforms, which included performance management (Curry et al. 2015, 22).

In sum, the scholarly research proposes that oversight should be positively associated with agency adoption of performance measures and use of performance information, but empirical research suggests its influence is limited and depends on what form the oversight takes, what functions it serves, how centralized or decentralized the overseen programs are, and the degree of autonomy that affected bureaucracies enjoy in the principal-agent relationship. A nuanced and synthesized view of the diverse literature leads to the proposition that if oversight exerts any positive influence on U.S. federal managers, it is more likely to lead agencies to adopt performance measures due to the need for agency compliance with legal reporting requirements and symbolic messaging to stakeholders to support agency legitimacy than it is to stimulate internal use of performance information for management decisions. In line with the argument of Andersen and Jakobsen (2018) that organizational innovation is more likely to be influenced by organizational factors than by external forces, this study proposes that federal manager use of performance information is more likely to be influenced by organizational conditions than by external oversight.

The study assumes the more positive the perceptions federal managers have of attention paid by oversight authorities to agency GPRAMA implementation, the more salient and influential the condition of oversight is to managers' decisions regarding its institutionalization. It models the construct of oversight as a condition in the system action situation and operationalizes it as a continuous latent factor variable that measures the shared variance of observable indicators of federal agency manager perceptions of attention paid to agency implementation of GPRAMA by the following oversight authorities: 1) OMB), 2) non-OMB

interagency organizations, 3) Congressional committees, 4) auditors, and 5) the public. Additional detail on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are proposed:

H_{2a} External oversight is positively associated with the adoption of performance measures in agencies.

H_{2b} External oversight is not associated with the use of performance information for management decisions in agencies.

We now shift from examining the system action situation to an analysis of organizational conditions in the organization action situation.

Conditions in the Organization Action Situation

Research into organizational conditions and their effects on the institutionalization of performance management has contributed critical theoretical and empirical perspectives that frame possible answers to the research questions. The scholarship on antecedents of performance measure adoption and manager use of performance information has identified several primary independent variables. These include: 1) goal clarity (or its inverse, goal ambiguity), 2) a cultural climate of accountability and empowerment, 3) credible leadership commitment, 4) availability of technical training, 5) agency capacity to support performance evaluations, and the 6) maturity of the agency's performance measurement system (Kroll 2015a).

This study embraces a longstanding scholarly tradition in the fields of organizational sociology and public management analyzing organizational buffering in the context of environmental changes and uncertainty on organizational performance (e.g. M. L. Lynn 2005; Meier and O'Toole 2008; Thompson 1967). A central claim, proposed by Thompson (1967), is that organizations balance responsiveness to the environment with the protection of core

technical functions from disturbances, and pursue buffering strategies to do this using three organizational layers with varying environmental exposures: "boundary-spanners" interacting with the environment, the protected technical core executing the organization's function, and managers between these layers controlling the diffusion of the environmental shocks into the technical core (M. L. Lynn 2005, 39). This study extends the postulate concerning agency performance to agency institutionalization of performance management practices, which should enable organizations to deliver improved performance in the future. It explores which combinations of conditions, in the context of a public agency's relationships with its external environment, positively influence managerial decisions that, while not necessarily intentional, have the effect of reducing buffering and permit or facilitate the diffusion of performance management practices into the technical core where they can become substantive, and not merely symbolic. An example would be a federal agency that extends performance management routines beyond the headquarters to operational line units where essential products or services are developed and delivered. Such an expansion would indicate deep institutionalization of performance management to the technical core, where performance information would be more likely to affect daily manager decisions, not merely satisfy official reporting requirements

In addition, as discussed previously, organizational theory predicts OI creates incentives for managers to be custodians and protectors of their agency's identity. As such, it can be considered as an organizational condition that may also buffer, or in some cases encourage, manager decisions to institutionalize performance management. Managers' efforts to institutionalize environmentally-imposed changes may challenge the workforce's consensus on an agency's identity and thereby generate resistance in some agencies and adaptation in others, depending on how they make sense of new meanings (McInnes et al. 2006, 3-4). As discussed

below, the public management literature has dedicated insufficient attention to the influence of agency identity on management decisions to institutionalize performance management reform. This study contributes original insights on this subtle relationship from the empirical context of the U.S. Government.

Goal Clarity

Goal setting and developing performance measures are elements of performance management systems which can be defined via "fiat, negotiations, or models" and their attainment can be used by managers to drive organizational incentives, budgetary decisions, and autonomy (Gerrish 2016, 50). Mockler (1984) claimed that agreement on program goals was a necessary condition for a PMS (as cited by National Academy of Public Administration 1999, 22). The selection of performance measures, nevertheless, ". . . tends to reflect the interests of those who comprise the dominant coalition within the firm" (Waggoner, Neely, and Kennerley 1999, 55). Regardless of internal political agendas and coalitions advocating for certain goals, long-established goal-setting theory in the psychology, public management, and organizational theory literatures holds that clear and specific organizational goals facilitate performance evaluation and feedback as part of a strategic planning process, and ambiguous goals hinder them (J. W. Lee, Rainey, and Chun 2009, 458; Locke and Latham 1990 as cited in Rainey 2014, 287; Sawyer 1992 as cited in Anderson and Stritch 2016, 212). Furthermore, the literature has established that goal clarity, along with the ability to measure performance based on attainment of organizational goals and the ability of organizational members to transform processes and predict likely results of alternative decisions, is a necessary condition for effective performance contractability, an important mediating variable on how performance management systems affect agency performance (Spekle and Verbeeten 2014 131-132). De Lancer Julnes and Holzer (2001)

predict that organizations with a goal-driven orientation are more likely to adopt and use performance measures (695). Empirical research generally supports these predictions. In a comparative transnational analysis, De Jong (2016) found the greater complexity of some U.S. agencies' policy goals, relative to those of counterpart Dutch agencies, was created by competing agendas of actors and stakeholders in the American political system and political appointees, and that this goal complexity complicated federal agency performance planning and reporting (210). In the research context of the US Government under GPRAMA, using GAO survey data from federal managers in 2007 and 2012-2013, Moynihan and Kroll (2016) identified efforts to overcome goal complexity through the use of priority goals: "…agency priority goals foster performance information use because they direct attention to a limited number of key objectives and indicators, thereby helping managers juggle conflicting interests and deal with information overload" (320).

The inverse of goal clarity is, of course, goal ambiguity. Chun and Rainey (2005a) 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" (2). The literature of political science and organizational theory has long held goal ambiguity to be widespread and systemically inherent in public organizations (Rainey and Jung 2010, 35). Public agencies may answer to multiple principals, who may give conflicting direction that creates goal conflict, a primary source of ambiguity (O'Toole and Meier 2015, 249). Rainey (2014) explains the effect of complex governance on manager incentives:

> Public organizations produce goods and services that are not exchanged in markets. Government auspices and oversight imposed on these organizations include such multiple, conflicting, and often intangible goals as. . . constitutional, competence, and responsiveness values. In addition, authorizing legislation often assigns vague missions to government agencies and provides vague guidance for public programs. Given such

> mandates, coupled with concerns over public opinion and public demands, agency managers feel pressured to balance conflicting, idealized goals" (152).

Goal ambiguity is a real concern for federal managers. It has been found to be positively associated with possibly undesirable political salience of agency goals to elected leaders and the media (Lee, Rainey, and Chun 2009, 458), and negatively associated with the creation of clear and outcome-focused performance measures (Rainey and Jung 2010, 36), performance information use (Moynihan and Pandey 2010, 855; Moynihan and Kroll 2016, 320), agency autonomy (Khademian 1995, 24), senior public manager engagement (Andrews and Mostafa 2019), and perceived organizational performance (Jung 2011; 2014; Chun and Rainey 2005b). Based on the literature, one may conclude that federal managers will generally perceive a positive relationship between the clarity of their agency's goals and the robustness of their agency's performance management system.

This study models the construct as a condition in the organization action situation and operationalizes it as a continuous latent factor variable that measures the shared variance of the observable indicia of agency manager perceptions of the extent to which 1) agency goals are prioritized, 2) agency goals align with program purposes, 3) management communicates goals to employees, and 4) managers are aware of agency goal achievement. Additional detail on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are proposed:

H_{3a} Agency goal clarity is positively associated with the adoption of performance measures in agencies.

H_{3b} Agency goal clarity is positively associated with use of performance information for management decisions in agencies.

A Cultural Climate of Accountability and Empowerment

The National Performance Management Advisory Commission (2010) observed that "performance management systems represent a fundamental change in organizational culture" (19). However, performance management systems may also represent a set of behavioral routines that develop within organizational cultures which value, or do not value, performance in the first place. Waggoner, Neely, and Kennerley (1999) argue that cultures which are risk averse and non-innovative hinder the institutionalization of performance measurement improvements (58). On the other hand, a results-oriented culture, i.e. one in which employees are mission-focused, has been shown to facilitate organizational learning (Moynihan and Landuyt 2009, 1099) which, as described previously, is essential to the three-step process of institutionalization of new practices. Measures of social norm that include public managers valuing the use of performance data to steer their organization were shown to influence use of performance information directly, and indirectly through manager intentions to use it (Kroll 2015, 209). As part of a study on the effects of employee training on performance management implementation in the U.S. Government, Kroll and Moynihan (2015) identified accountability as a consistently significant and substantial control variable (416). A results-oriented culture has multiple values traditionally associated with it. In this study, the two primary ones under investigation are a shared sense of accountability and manager empowerment.

Organizations with a cultural climate that supports performance management are more likely to elicit discretionary efforts by employees to participate in performance management business processes (Moynihan 2013, 23-24). Culture, however, encompasses a rich tapestry of social perspectives, norms, and values across diverse groups, so variance in the preponderance of performance-related norms likely affect agency responses to performance management reforms. Certain orientations and norms in organizational cultures have been found to be associated with

the adoption of performance measures and the use of performance information in government agencies. For example, to support GPRAMA implementation, OMB Director Jack Lew called for cultural changes across federal agencies: "Critical to success is a culture where agencies constantly ask, and try to answer, questions that help them find, sustain, and spread effective and efficient programs, practices, and policies that support the agency's mission" (U.S. Office of Management and Budget 2011, 1). Destler (2016) notes several leading scholars of organizational change have identified a cultural climate where employees feel supported, psychologically safe, free to experiment, and express their opinions builds trust and psychological readiness to evaluate, integrate, and eventually accept new reform practices, which helps to embed performance management institutionalization within frontline units where path dependency in operational practices is greatest (204).

Performance management philosophy advocates that organizations seek to achieve their goals, so a goal-seeking cultural orientation that tolerates manageable risk taking is often cited as a necessary component of agency culture. De Lancer Julnes and Holzer (2001) found that a goal-seeking orientation in agency culture is positively associated with the adoption of performance measures, but not significantly associated with the use of performance information when controlling for adoption in their model. On the other hand, their analysis found that a risk-taking orientation was not significant when controlling for strong "rational/technocratic" factors like external and internal requirements, resources, information, and goal orientation (698-701). This suggests the existence of high levels of risk aversion in public bureaucracies leads to statistically-inconsequential levels of risk taking as a rule. Nitzl, Sicilia, and Steccolini (2019) proposed that a results-focused orientation with norms of productivity, goal seeking, and efficiency; a developmental orientation with norms of flexibility and innovation; and a group

orientation with participatory norms such as member empowerment and a common vision are positively associated with an organizational culture oriented to the NPM philosophy (691-692).

Despite contrary expectations of NPM theory, which emphasizes trust and delegation of authority by political principals to agency leaders in return for agency transparency about its performance, an orientation of mission accountability in an agency's culture may positively predict performance information use (Doering 2012; Moynihan and Kroll 2016, 319), but not when employees fear performance accountability will be used in a punitive way against them (Arnaboldi, Lapsley, and Steccolini 2015; Moynihan 2013).[13] Employee recognition is also a tool for public managers that contributes to a larger cycle of incentives, actions and consequences to shape organizational culture in a direction favorable to improved performance and adoption of performance management systems (Hartmann and Khademian 2010, 849-852). Austin and Ciaassen (2008) and Fernandez and Moldogaziev (2011) found collaboration in the form of "open and honest dialogue" to be an important participatory value that favors performance management reforms (as cited by Destler (2016), 204). Similarly, Poole (2019) found a cultural orientation towards openness, treated here as a willingness to collaborate, predicts higher quality and timeliness in agency responses to requests for information (224).

Distilling these various claims from these multiple lines of theory on performance-oriented cultural norms, this study models the construct of a cultural climate of accountability and empowerment two ways. For the quantitative analysis, it models it as a latent factor variable in the organization action situation that measures the shared variance of the observable indicia of

[13] Furthermore, Metzenbaum (2013) advises "…a direct linkage between target attainment and incentives often plays out in practice with a retreat to timid targets, compromising the innovation-inspiring value of a stretch goal, or measurement manipulation, necessitating more spending on monitoring" (858).

agency manager perceptions of the extent to which 1) managers are empowered with the authority they need to accomplish strategic goals, 2) managers are held accountable for agency strategic goal accomplishment, 3) managers are held accountable for the programs they are responsible for, 4) employees receive recognition for goal accomplishment, 5) top leaders demonstrate a commitment to achieving results, and 6) managers use performance information to collaborate with others on effective approaches. For the qualitative analysis, it models it as a condition in the organization action situation that is operationalized as a fuzzy set. In this set, agencies are outside, partially in, or fully inside the set of agencies with cultures of strong accountability and manager empowerment. Additional detail on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are offered:

H_{4a} An agency culture of accountability and empowerment is positively associated with the adoption of performance measures in agencies.

H_{4b} An agency culture of accountability and empowerment is positively associated with the use of performance information in agencies.

Credible Leadership Commitment

According to Fernandez and Rainey (2006), the literature on organizational change has found considerable evidence that ". . . top-management support and commitment play an essential role in successful change in the public sector. . ." and that ". . .top-management for support for change often requires the cooperation of top-level career civil servants in addition to politically-appointed executives" (171). Van Wart (2003) reviewed the literature on public leadership and found a strong association between public leadership and public sector reform with a clustering of views between what he refers to as the "entrepreneurial" and "stewardship" communities (219).

Turning to performance management reform specifically, agency senior leadership support has been empirically shown to have a highly significant and substantial correlation with long-term benefits of using performance measures, at least at the local level (Melkers and Willoughby 2005, 187-188). Favorable federal manager perceptions of political support may encourage them to overcome the organizational challenges of institutionalizing reform in their agency (Franklin and Long 2003 as cited by K. Yang and Pandey 2008, 342). Accordingly, the credible commitment of public sector leaders has been postulated and shown to be a positive predictor of different forms of performance management reform policy institutionalization. Dull (2009) explains ". . .credibility is a necessary condition for performance management. If managers fear that leadership is not credible—if they suspect leaders lack competence, authority, or trust, the capacity to create credible information will not take hold" (257). An expected response by managers who perceive credible commitment by senior leaders is their investment of the "cognitive resources necessary to alter routine performance" (Dull 2009, 260). Because leader and manager attention is always limited and must be focused on one issue at a time (Jones and Baumgartner 2005, 45), devotion of their valuable attention to performance management at the opportunity cost of devoting it to other organizational priorities should be perceived by other managers as an indirect but visible indication of their credible commitment to performance management reform.

Given the importance of managerial attention, is attention to performance management reform, by itself, sufficient for a federal manager to be seen by subordinate managers as "credibly committed?" This would seem unlikely in the absence of positive action to institutionalize new processes. The literature presents two schools of thought about which style of public leadership is most suitable: transformational (or entrepreneurial) leadership, which is

often associated with NPM philosophy, and integrative leadership, which also seeks to improve organizational and management systems but more keenly respects the governance limitations under which public managers must function (Moynihan and Ingraham 2004, 428). The literature on transformational public sector leaders claims they can heighten managers' public service motivation and thereby increase their use of performance information (Kroll and Vogel 2014, 3). They can also communicate a vision, effectively lead top managers, serve as a role model, and change organizational climate, strategies, and business processes (Gang, Wang et al. 2011, 233-251).

Regardless of their transformational identity, all public leaders can exert direct influence on their agencies through a set of mechanisms, including strategic planning, mobilizing external political support, acquiring resources, reporting performance to their "authorizing environment" to improve their own authority to act with discretion and initiative, publicizing initiatives, fostering innovation and changing manager attitudes regarding management reforms (Moore 1995, 130, 282-283). Supporting these claims empirically, Moynihan and Hawes (2012) found management interaction with the oversight environment to be positively correlated with performance information use in Texas school districts (S100-S101). Their findings complemented prior research by O'Toole and Meier (2011), who found networking with the oversight environment by school district superintendents in Texas substantially improved student performance on standardized tests of schools in their districts (72-73). Entrepreneurial public managers have been found to be positively correlated with citizen satisfaction with English local government efficiency, responsiveness, equity, and effectiveness (Andrews and Van de Walle 2013), suggesting further research would show they can influence the adoption of performance measures and use of performance information within their agencies as well.

Scholars of integrative leadership are more modest in their claims about the level of discretion public sector managers enjoy. Rather than bypassing existing administrative business process, integrative managers work through financial management, human resources, capital management, and information technology sub-systems to allocate organizational capacity to enhance performance (Moynihan and Ingraham 2004, 428-429). Moynihan, Pandey, and Wright (2012) argue even transformational leaders must use standard management practices to institutionalize reform, and can succeed not only through direct action but also by "setting the table" for greater management use of performance information indirectly by clarifying goals and developing a performance culture (159). Van Wart (2013) takes the claim a step further, arguing that the institutionalization of performance management has increased interest in transactional leadership (525). There is, thus, plenty of room for many styles of leaders and leadership to be credible and effective drivers of organizational change, including the adoption and use of performance management practices.

This study models the construct of credible leadership commitment two ways. For the quantitative analysis, it models the construct as a condition in the organization action situation and operationalizes it as a continuous latent factor variable. The variable measures the shared variance of the observable indicia of agency manager perceptions of agency senior leadership demonstrating a strong commitment to using performance information to guide decision making, and agency managers/supervisors aligning program performance measures with agency-wide goals and objectives. For the qualitative analysis, it models it as a condition in the organizational action situation and operationalizes it as a fuzzy set in which agencies are outside, partially in, or fully inside the set of agencies with senior leaders whose commitment to performance management is perceived as highly credible by federal managers. Additional detail

on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are advanced:

H_{5a} Credible leadership commitment is positively associated with the adoption of performance measures in agencies.

H_{5b} Credible leadership commitment is positively associated with the use of performance information in agencies.

Performance Management Training

The effect of employee training on performance management institutionalization has been only lightly treated in the literature, and why it matters is poorly understood (Kroll and Moynihan 2015, 411-412). Even the broader topic of organizational learning in the public sector has enjoyed little coverage (Moynihan and Landuyt 2009, 1097). Kroll and Moynihan (2015) proposed training may inform employees of requirements, establish the reform as a new norm, or build additional organization capacity for implementation (413). They found that performance management training increased the attention of managers on performance information and strategic goals, but failed to equip them with skills to deal with difficult measurement problems (418). Based on employee and parent surveys and third party reports, recent quantitative research into over 1,000 New York City schools found training climate was positively associated with the performance management function of organizational strategy development, but not performance data use or continuous improvement (Destler 2017, 520-528). A possible explanation for the latter findings may be found not in the merits of the offered training program, rather in the quality of the performance data available to employees. Employees interpret and integrate their experiences in learning forums, but their faith in the available performance information depends on the credibility of the performance data and their ability to assess it

(Deschamps and Mattijs 2018, 488). Kroll (2015) identified learning forums as a "promising" impact factor for performance information use (474). Thus, the influence of training in performance management is not yet fully understood in the context of confounding variables, and considerable foundational research is needed, but early research is promising.

This study models the construct of training in performance management as a condition in the organization action situation. For the quantitative analysis, it operationalizes the construct as a continuous latent factor variable that measures the shared variance of the observable indicia of agency manager perceptions of the extent to which the agency offers training in 1) strategic planning, 2) setting program performance goals, 3) developing program performance measures, 4) assessing the quality of performance data, 5) using program performance information to make decisions, and 6) methods to link program performance with the achievement of agency goals. Additional detail on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are proposed:

H_{6a} Performance management training is positively associated with the adoption of performance measures in agencies.

H_{6b} Performance management training is positively associated with the use of performance information in agencies.

The Capacity to Evaluate Performance

Agency capacity to conduct performance evaluations is cited often in the literature as a predictor of different forms of performance management institutionalization (e.g. Kroll 2015b; Moynihan and Hawes 2012; Yang and Hsieh 2007). What may seem intuitively obvious to some, organizations require the capacity to conduct performance measurement in order to institutionalize a system of performance management (Lee and Clerkin 2017, 608). Moynihan

and Kroll (2016) identify staff, technology, and quality data as required organizational capacities for data-driven performance reviews (317). The concept of capacity in public organizations includes officials' education and professional credentials (Fukuyama 2013, 354). Therefore, staff with performance management-related knowledge and skills should be counted as part of capacity for performance evaluation. Based on Punctuated Equilibrium theory, adding capacity for the evaluation of performance information should theoretically mitigate constraints on managers' attention (Epp and Baumgartner 2017, 248) and thereby give them, or their staffs, more time and opportunity to adopt performance measures and use performance information. To implement laws in a manner that achieves their intended purposes, Fernandez and Rainey (2006) argue that agencies require "ample funding…to provide them with the administrative and technical capacity" and note that some public management literature has found that a lack of resources may impede organizational change (172). How resources are allocated within the organization to build capacity for reform implementation appears to matter as well. Reform implementation is likely to be moderately successful if modest levels of resources are distributed more or less equally over time to support multiple reform changes because it minimizes policy competition, vice the opposite scenario in which implementation costs are allocated unevenly and result in resource competition between elements that favors less expensive initiatives, unless there is substantial commitment to the higher cost implementation plan by senior leaders (Chackerian and Mavima 2001, 373-374).

This study models the construct of performance evaluation capacity as a condition of the organization action situation and operationalizes it two ways. For the quantitative analysis, it operationalizes the construct as a continuous latent factor variable that measures the shared variance of the observable indicia of agency manager perceptions of the extent to which the

agency 1) has sufficient analytical tools for managers to collect, analyze, and use performance information, 2) has sufficient staff with the knowledge and skills needed to analyze performance information, 3) is investing the resources needed to ensure that its performance data is of sufficient quality, 4) is investing in resources to improve the agency's capacity to use performance information. For the qualitative analysis, it models it as a fuzzy set, in which agencies are outside, partially in, or fully inside the set of agencies with adequate capacity for performance evaluation. Additional detail on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are advanced:

H_{7a} The capacity for performance evaluation is positively associated with the adoption of performance measures in agencies.

H_{7b} The capacity for performance evaluation is positively associated with the use of performance information in agencies.

The Maturity of the Performance Measurement System

The maturity of an agency's performance measurement system in public organizations is also cited as a predictor for performance management institutionalization. The author has been unable to identify a definition for the maturity of performance measurement practices in the public management literature, but Bititci et al. (2015) propose a definition from the field of production economics that seems to conform with sociological neo-institutionalist theory:

> . . .the ability to respond to the environment in an appropriate manner through [performance measurement and management] practices. This response is generally learned rather than instinctive, but maturity. . .is a reflection of the appropriateness of its measurement and management practices in the context of its strategic objectives and in response to environmental change. (3065)

Their model of maturity envisions a spectrum of development through five stages: "ad hoc," "basic," "defined," "managed," and "optimizing" (3066). Similar to the three phases of

institutionalization proposed by Tolbert and Zucker (1996), the phases of maturation of a performance measurement system are differentiated based on ". . . the level of formality, sophistication and embeddedness of practices" (3065). In other words, the depth of their institutionalization defines the maturity of performance measurement practices.

Proeller, Kroll, and Meier (2012) propose the maturity of performance measures in an agency predicts increased performance information use. Kroll (2015) confirms the literature has identified measurement system maturity as an "important driver" of performance data use and proposes indicia for a mature measurement system, including reports that satisfy requirements, the use of benchmarks, accessible performance measures, achievable performance goals, satisfactory availability of performance information that links to goals and strategic plans, and data diversity (471-472). In this study, the maturity of an agency's performance measurement system, which intuitively seems like a dependent variable or an outcome of the institutionalization process, is treated as a predictor variable and condition for the institutionalization of a performance management system because 1) performance measurement is a component and necessary precursor of a performance management system (van Dooren 2005, 363), and 2) it can be treated as an autoregressive control variable because it represents the prior institutionalization of performance measurement practices in an agency. Including it as a control variable for all other institutionalization indicia in the inferential model allows between-agency comparisons of all other system action situation and organization action situation conditions in this cross-sectional study that are not biased by marginal agency advantages from prior investments in performance measurement capacity.

This study models the construct of performance measurement system maturity as a condition of the organization action situation and operationalizes it two ways. For the

quantitative analysis, it operationalizes the construct as a continuous latent factor variable that measures the shared variance of the observable indicia of agency manager perceptions of the extent to which the agency 1) determines meaningful measures, 2) standardizes definitions for measures of performance, 3) can obtain valid and reliable data, and 4) is able to obtain data in time to be useful for decision making. For the qualitative analysis, it operationalizes it as a fuzzy set in which agencies are outside, partially in, or fully inside the set of agencies with mature performance measurement systems. Additional detail on the data source and measurement strategy is provided in chapter 4. Two testable hypotheses are proposed:

H_{8a} The maturity of a performance measurement system is positively associated with the adoption of performance measures in agencies.

H_{8b} The maturity of a performance measurement system is positively associated with the use of performance information in agencies.

Agency Identity and Bureaucratic Type

The final organizational factors proposed in the model are "agency identity" and "bureaucratic type." Based on the previous discussion on organizational identity (OI), this study will employ the equivalent term "agency identity" instead and define it more parsimoniously than OI as the basic organizational characteristics defined in law and policy that govern a public bureaucracy's mission mandate and its level of autonomy, hence its core mission identity. While acknowledging the claims of Carpenter and Krause (2012) and Wilson (1989) that reforms may change both the functional tasks and level of autonomy of agencies over time (Nielsen and Baekgaard 2015, 548), this study argues that the existing tasks and autonomy of agencies also have consequences for their institutionalization of reforms. As Jennings and Haist (2004) argue, it may also determine whether agency use of performance measurement is substantive or merely

symbolic (182). Verhoest and Wynan (2018) aptly articulate the core theoretical claim: "…task characteristics and the related technical environment affect organizational practice, and thus the adoption and use of performance management techniques" (626). The degree of task measurability depends on the kinds of tasks an organization is chartered to perform (Wilson 1989) and is postulated to be positively associated with the institutionalization of an organization's performance management system (Van Dooren 2005).

To be effective, public sector reformers should clearly understand what is being done in affected bureaucracies (Wilson 1989, 11-12). Agency tasks can be complex or routine, and the public management literature, echoing predictions from OI theory, has shown public agencies with more routine tasks, on average, have clearer goals (Rainey and Jung 2010, 42-43) and perform relatively better than agencies which must grapple with ambiguity in their governing policies (Gormley and Balla 2008, 203). Clearer goals, as described previously, often reflect clearer OI and, so long as organizational routines are not incommensurate with performance management reforms, should enable agencies to adopt and use more observable (and objective) performance measures and performance information, and thereby be able to make a claim to good performance. In contrast, some critics argue that managers working in a performance management regime may neglect organizational goals which are difficult to measure (Boyne 2010, 210; Radin 2006, 43).

Task measurability therefore provides a tractable basis for differentiating public organizations, and offers a useful way to express an agency's overall orientation towards performance. An agency that performs functions which are objectively measurable, or whose outcomes are measurable and to a degree controllable, should be more likely to adopt and use performance measures, and that usage should have a greater relative effect on performance than

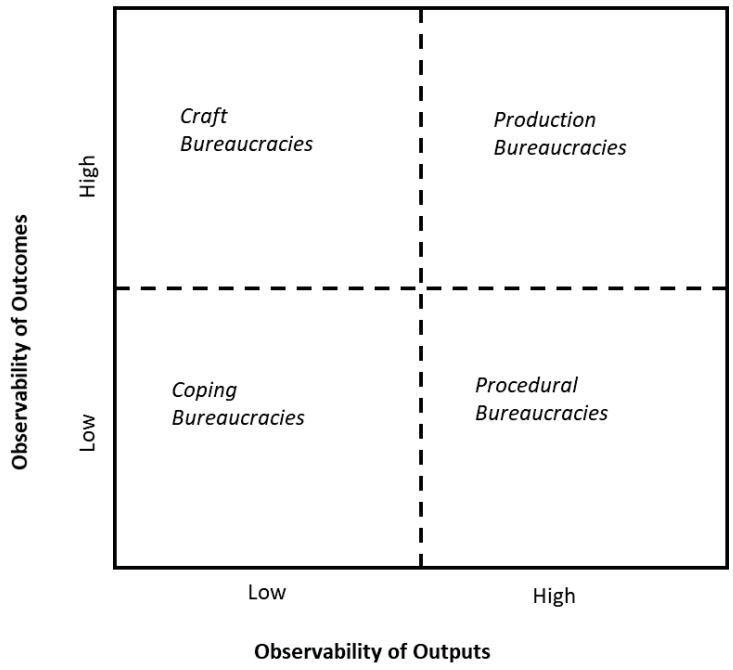
it would in agencies whose functions are difficult to observe (Jennings and Haist 2004, 184-185). This study operationalizes the construct of agency identity as a measure of performance observability in a manner that permits between-agency analysis in two ways. First, in the quantitative analysis, it models agency identity as a condition in the organization action situation and operationalizes it as a continuous latent variable, at level two of a hierarchical linear model, that measures agency performance orientation (i.e., the perceived measurability of agency outputs and outcomes) while controlling for the effects of all other organizational condition predictors in the inferential model, described above. Additional detail on the data source and measurement strategy is provided in chapter 5. Two hypotheses are proposed:

$H_{9a.}$ A significant relationship exists between an agency's identity and its adoption of performance measures.

$H_{9b.}$ A significant relationship exists between an agency's identity and its use of performance information.

In chapter 5, agency identity is also analyzed by agency bureaucratic type, employing the taxonomy of bureaucratic types proposed by J. Q. Wilson (1989). This taxonomy holds that public agencies of similar orientation towards performance can be clustered into bureaucratic groups that roughly share the same degree of measurability of their organizational outputs and outcomes. Wilson observed that the relative level of observability of performance outputs and outcomes could help to explain variance in agency adoption of performance measures and their use of performance information, and proposed a taxonomy (figure 2.2) that categorized all public bureaucracies into four ideal-type categories: Production, Procedural, Craft, and Coping (Wilson 1989, 158-171).

Figure 2.2. James Q. Wilson's Typology of Bureaucracies (1989)



Source: Author's own illustration.

In Production bureaucracies, outcomes and outputs are easily observable, so they are more likely to adopt performance measures and use performance information to the greatest extent to evaluate their outputs and outcomes. The use of performance measurement should also have the greatest impact on performance in Production bureaucracies (Jennings and Haist 2004, 185). Procedural agencies have observable outputs but not observable outcomes, so they are expected to make greater use of output-type measures and use performance information for those. These agencies rely heavily on efficiency measures (Jennings and Haist 2004, 185). Craft organizations have observable outcomes but not outputs, so they should adopt and make use of outcome measures but make relatively less use of output measures, and performance information use should align primarily to measuring agency outcomes. Any use of output measures will likely be merely symbolic (Jennings and Haist 2004, 185). Coping organizations have few observable outcomes or outputs and so are predicted to make little use of either outcome or

output measures, or any performance information. As a result, their measures may be indirect and often unclear, or expressed at the business unit level (Jennings and Haist 2004, 185).

This study models the construct of "bureaucratic type" using Wilson's taxonomy of bureaucracies and operationalizes it as a nominal categorical variable that identifies agency alignment to bureaucratic category based on an agency's performance orientation relative to other federal agencies according to federal manager perceptions of their agency's use of output and outcome measures. Additional detail on the data source and measurement strategy is provided in chapter 5. Based on Wilson's theoretical predictions, this study advances four testable hypotheses:

$H_{10a.}$ Production agencies are more likely to adopt output measures than Craft and Coping agencies are, and are more likely to adopt outcome measures than Procedural and Coping agencies are. They are more likely to use performance information than all other bureaucratic types.

$H_{10b.}$ Procedural agencies are more likely to adopt output measures than Craft and Coping agencies, and are less likely to adopt outcome measures than Craft and Production agencies. They are more likely to use performance information than Coping agencies are.

$H_{10c.}$ Craft agencies are more likely to adopt outcome measures than Procedural and Coping agencies are, and are less likely to adopt output measures than Production and Procedural agencies are. They are more likely to use performance information than Coping agencies are.

$H_{10d.}$ Coping agencies are less likely to adopt performance measures of any kind and to use performance information for any purpose than Production, Procedural, and Craft agencies.

Now we transition to the third stage of the inferential framework and focus on how the literature predicts public managers respond to these institutional and organizational influences as they make decisions to institutionalize performance management reform.

Institutionalization Decisions in the Individual Action Situation

This final portion of the inferential framework at figure 2.1 in the individual action situation is inspired by the methodological approach of van Dooren (2005) to facilitate the evaluation of public manager decisions to institutionalize performance management along two primary dimensions of praxis: the adoption of a diverse set of performance measures, and the use of performance information generated by those measures for management decisions (367).[14] This study treats these practices as the measurable outcomes of the performance management reform policy institutionalization process. This study conceptualizes public managers as agents of institutionalization in the way Simon (1997) does as boundedly-rational with limited attention who make "satisficing" decisions based on heuristics and using a subset of potentially available information they perceive as relevant (118-119). A critical assumption, based on sociological neo-institutionalism, is that there is no basis to assume managers will be perfectly rational, in the sense of consciously weighing all possible alternatives and making precise cost-benefit calculations, in their decisions to institutionalize GPRAMA. For this reason, their decisions to adopt specific performance measures, and whether and how they make evidence-informed decisions using performance information, are highly complex and challenging to model.

The use of detailed structured psychological or sociological profiles for study participants would be useful and appropriate for measuring the effects of attitudes and biases on manager

[14] Van Dooren (2005) credits his methodology, in turn, to the conceptual distinction Beyer and Trice (1982) and de Lancer Julnes and Holzer (2001) had made between the adoption and implementation of policy.

responses to incentive structures, but is beyond the scope of this study.[15] Yet, public leaders' attitudes about organizational reinvention were shown to be a significant predictor of variance in agency adoption and implementation of performance management practices (Brudney, Hebert, and Wright 1999; Kroll 2014). In the individual action situation, the escalation of a performance management regime can lead to competing institutional logics. For example, an imposed logic of performance-based accountability can bring a public manager accustomed to a professional ethic and mindset into role conflict, resulting in disagreements with performance systems managers over what aspects of performance should be measured and how (Pollitt 2013, 353; Pfiffner 2019, 406). Beyond technical disagreements over measurement procedure, a larger conflict in values and role expectations for public officials may occur when efforts to impose a new culture of managerial entrepreneurialism in an agency run into conflict with traditional public expectations of public employees to adhere to democratic standards of public sector behavior (du Gay 1996, 153).

The willingness of agency performance management system team members and operating unit managers to implement GPRAMA policy requirements, just as with the performance of government employees at street level,[16] can also be influenced by a wide range of personal cognitive factors. Drawing from the literature of employee behavioral theories, Thomann (2015) identifies several examples: feelings of powerlessness, perceptions of the societal or client meaninglessness of the policy, divided responsiveness to multiple professional

[15] For an overview of this literature, see Kroll (2015a).

[16] In this study, action officer level federal employees (i.e., non-senior executives) who are responsible for establishing performance management processes and embedding related practices into daily work life across their organizations (even those working at the level of the departmental or agency headquarters) are considered to be "street level," because they work at the exact point of policy implementation. Their actions are necessary, but not necessarily sufficient, for GPRAMA institutionalization in federal agencies.

and societal accountabilities outside of formal hierarchical structures, individual professional or ethical values which may conflict with the values or practices associated with the policy, or in some cases perceptions of role conflicts between policies and clients (181-182). Therefore, using the research capacity available, this study investigates how federal agency performance system manager attitudes towards performance management, their intrinsic public service motivation, and their perspectives on the system and organization action situation conditions affect their willingness to advocate for the institutionalization of GPRAMA practices. The research design, results, and findings are presented in chapter 4.

Adoption of Performance Measures

The literature has come to define performance information as "…feedback information on outputs and outcomes of the public service as well as its efficiency (a ratio of outputs and inputs) and its effectiveness (a ratio of outcomes and outputs)" (Kroll 2015b, 461). A meta-analysis of the public management literature by Andrews, Boyne, and Walker (2011) revealed other measures of performance have been of research interest in the field, such as equity, responsiveness, and trust (235). Some scholars have proposed government agencies adopt democratically- or socially-oriented measures, such as social justice or accountability (Denhart and Denhart 2015, 184; Radin 2006, 214-231; Van Thiel and Leeuw 2002, 277). To maintain parsimony and to support analysis of the effects of agency identity and Wilson bureaucratic type, this study focuses primarily on outcome and output measures, but will also include empirical evidence concerning agency adoption of measures of efficiency, customer satisfaction, and quality to support the interpretation of the results for the measurement of the dependent variable concerning the institutionalization of performance measure adoption in the inferential model.

Which measures managers adopt in practice is abased on a complex and not fully rational decision making process designed to achieve optimal organizational or program performance outcomes. What psychological factors condition their responses to incentives? Public managers are biased by a wide range of cognitive limitations (Belle, Cantarelli, and Belardinelli 2018, 835-836), swayed by interest group influence (Wildavsky and Caiden 2004, 192), contend with more severe goal ambiguity than private managers must (Jung 2011, 193), experience only weak extrinsic incentives to implement performance management (Moynihan and Lavertu 2012, 594), and, just as with citizens, are subject to framing effects when interpreting performance information (Fuenzalida, Van Ryzin, and Olsen 2020, 13). The influence of material, solidarity, and purposeful incentives on management innovation in the public sector has, nonetheless, been empirically demonstrated (Hopkins 2016). Yet, in the face of conflicting demands on their time and attention, and external and internal agendas for their organization or for themselves, incentive structures may not always be strong enough to induce managers to follow through and adopt some types of performance measures, at least not to the degree intended by legislators.

While acknowledging the importance of individual personality traits and value preferences, this study argues that manager decisions to adopt performance measures will depend greatly on the configuration of the systemic and organizational conditions they face. Spekle and Verbeeten (2014) suggest public managers seeking higher performance from their agencies will prefer measures that "…are consistent with the character of [the agency's] activities" (132). Julnes and Holzer (2001) argue public managers may choose to act symbolically by adopting performance measures merely to satisfy legal or regulatory requirements, but not fully implement the reform in practice because "rational/technocratic" factors will tend to influence

the adoption of performance measures, whereas "political/cultural" factors will tend to influence their implementation (695-698).

The literature seems to have reached a consensus on the claim that the validity, reliability, quality, availability, timeliness, standardization, and meaningfulness of performance data, all indicators of performance measurement system maturity, affect manager adoption and acceptance of performance measures. Rainey (2014) cites conceptual challenges in measuring some kinds of performance (451) and notes "…measuring goal achievement raises challenges for organizations of all types, and especially public and nonprofit organizations" (475). Common assumptions among some analysts concerning information are actually empirical questions, such as whether or not it is easily available, it is neutral, the measured phenomenon is understood, cause and effect relationships can be defined, baseline information for comparison is available, and all forms of performance are measurable and quantifiable (Radin 2006, 184-185). Spekle and Verbeeten (2014), citing Holstrom (1982) and Roberts (1984), observe that performance metrics may be "noisy"[17] and create potential conflicts of interest for managers:

> . . .it is often the case that realized performance as it gets measured depends not only on the individual manager's efforts and action choices, but also on decisions taken elsewhere in the organizations, on the cooperation of external parties, and on uncertain, uncontrollable events. In these situations, performance measures become noisy and do not adequately reflect managerial achievements. Noisy metrics are less informative about managerial performance, and may impose undue risk on the manager if his or her compensation or career prospects are based on these performance measures (136).

A particular challenge for public agencies in analyzing performance information cited in the literature is the difficulty of measuring outcomes, or program effects, on society. Trying to measure program outcomes that can take years or decades to achieve, as is the case with many

[17] "Noisy" metrics mean there are sources of random error in measurement from confounding variables that are difficult to analytically separate from the phenomenon being measured, i.e., managerial achievements, business strategies, operational processes, etc.

social programs, may be a daunting endeavor for many agencies, leading them to avoid analyzing outcome data at all and to favor other types of goals that are more measurable in the short run such as output measures (Lee and Clerkin 2017, 605).

The theoretical expectation here is that agencies which have adopted a diverse set of performance measures by type should be considered as advanced in this dimension of institutionalization. This expectation is grounded on the normative view of the GAO that ". . . high-performing, results-oriented organizations establish a set of measures to gauge progress over various dimensions of performance" (U.S. General Accounting Office 2001, 21). All U.S. federal agencies manage a diverse set of programs with a wide array of potentially measurable performance characteristics. Developing innovative methods to measure performance which overcome conceptual difficulties as noted in the literature indicates deeper institutionalization of performance measures. Agencies which have been able to transcend traditional practices of measuring internal activities and program outputs to measure program outcomes and customer satisfaction, for example, have institutionalized performance measurement more fully.

This study therefore models the construct of adoption of performance measures as a condition in the individual action situation and operationalizes it two ways. For the quantitative analysis, it operationalizes the construct as a continuous latent factor variable that measures the shared variance of the observable indicia of agency manager perceptions of the extent to which an agency has adopted output, efficiency, customer service, quality, and outcome measures. For the qualitative analysis, it operationalizes it as a fuzzy set outcome in which agencies are outside, partially in, or fully inside the set of agencies with a diversity of performance measures in use, employing the same observable indicia. More information on the data source and measurement strategies is in chapters 4 and 5.

Use of Performance Information for Management Decisions

Use of performance information for specific management decisions is the final condition of the inferential model. A robust literature focused on this subject has blossomed since Moynihan and Pandey (2010) advocated for performance information use to become a construct for research in the study of governance (859-862) and much of it has already been introduced above associated with the organizational conditions. When institutionalized using new routines for measurement and reporting, the literature shows the use of performance information becomes more likely ". . .when ensuring a certain quality, relevance, and regularity of the produced information" (Kroll 2015a, 202). Holm (2018) argues much of the performance management literature lacks empirical evidence of functional performance management practices (303), but finds that information about negative performance may lead public managers to increase the priority for agency goals offering the highest return on investment (315). Supporting evidence for Holm's latter claim is seen in a 2017 GAO study which evaluated U.S. federal agency progress in implementing GPRAMA. This analysis measured how well public managers perceive agencies use performance information to develop program strategies, set program priorities, allocate resources, identify program problems, take corrective actions, change work processes, set performance goals, manage contracts, and reduce duplication across programs (Government Accountability Office 2017, 35). In the context of GPRAMA, these uses of performance information apply to operational and exploratory uses, but rarely to incentive uses (for more on these, see Spekle and Verbeeten 2014, 134-135). This alignment is both fortuitous and fortunate as it avoids possible commingling of incentives from the organizational and system action predictors with any potential incentives from the agency's performance management system, which could lead to endogeneity in the model.

This study models the construct of performance information use as a condition in the individual action situation and operationalizes it two ways. For the quantitative analysis, it operationalizes the construct as a continuous latent factor variable. This variable measures the shared variance of the observable indicia of agency manager perceptions of the extent to which managers make use of performance information for a range of management decisions. Based on previous developmental work by the GAO and used in periodic GAO surveys of federal managers, a large and representative sample of such management decisions includes: 1) developing program strategies, 2) setting program priorities, 3) allocating resources, 4) identifying problems, 5) taking corrective actions, 6) adapting new program approaches or changing work processes, 7) setting or revising performance goals, 8) developing and managing contracts, and 9) streamlining programs to eliminate duplication. For the qualitative analysis, the study operationalizes performance information use as a fuzzy set outcome, in which agencies are outside, partially in, or fully inside the set of agencies using PI for a wide variety of management uses, employing the same observable indicia.

Before proceeding into the data sources, research designs, and study findings in chapters 4 and 5, chapter 3 presents a comprehensive historical backdrop through the analytic lens of historical institutionalism to add both a longitudinal perspective and contextual insights about the influence of the most significant system action and organizational conditions in the inferential model during the long historical arc of the modern era of performance management reforms in the U.S. Government which began in the early 1990s.

Chapter 3. The History of Performance Management Reform in the U.S. Government

According to Williams (2002, 457), the roots of performance-related reforms in American government have a long history going back to 1906, during the Progressive Era, at the New York Bureau of Municipal Research. During that era, he indicates public sector performance measures focused on accounting and budgeting, work performance, and social needs and outcomes, and were an important element of the reformers' agenda to grow government without requiring more tax revenue (458). Schachter (2002) observed that, in order to mobilize and sustain elite support for their cause, the Progressives framed their campaign as a movement to achieve greater "efficiency" by adapting Frederick Taylor's Scientific Management to the public sector context:

> To succeed at enlarging the administrative state, all reformers had to use a vocabulary that would influence legislators and judges . . . Presenting facts that one could claim were supported by science and conducive to efficiency afforded. . .reformers a way to hand members of the political elite a reason to sanction change. It gave decision makers who voted for reforms a way to rationalize their response to other elite members. (573)

Thus began a multi-generation campaign for good government, with earnest reformers seeking to improve government to serve the public good, while strategically employing useful, if overly simplified, rhetoric to sustain popular and political support. While the nation has seen several "tides" of government reform since then focused on oversight and waste reduction, the most recent one, which Light (1997) dubs "Liberation Management," aimed at fostering results-oriented management and improved performance, began in the early 1990s (Light 1997). A central assumption of this study is that this has been long enough to have allowed all stakeholders and participants to complete most actions necessary and sufficient to achieve full institutionalization of performance management reforms. If so, then cases of partial institutionalization in agencies can be considered as *prima facie* evidence of the moderating

effects of system action level and organizational conditions, described in the literature review in chapter 2, on agency-level decisions to embed performance management reforms in daily practice. But which conditions matter the most for agency decisions on institutionalization? Does their presence, or absence, generate incentives or disincentives that encourage (or discourage) federal manager decisions to adopt performance measures or to make use of performance information for management decisions? As introduced in chapter 1, this dissertation seeks to explore these questions. This chapter starts that investigation by examining what the modern history of performance management reform in the U.S. Government reveals about the effects of the most important conditions on the institutionalization of reform practices.

Theoretical claims in the literature, discussed in chapter 2, suggest that in most cases the presence of the conditions in this study's inferential framework generate positive incentives for federal managers. In some cases, however, the evidence suggests the presence of certain conditions, such as political influences, may create risk-based disincentives as well, or that their absence may encourage reform-minded entrepreneurial managers to innovate. Because conditions are likely to matter to reform institutionalization in a pattern of interaction with other conditions, evaluating their significance for GPRAMA institutionalization through the lens of historical institutionalism would ideally require a historical analysis of the dynamic relationships between all conditions in the inferential model and how they affected reform policy institutionalization over decades – a daunting task beyond the scope of the current research effort. In lieu of such a difficult endeavor, the following historical review and critical analysis of the origins and evolution of the performance management reform movement has more modest aspirations to create a contextually-grounded foundation for understanding the dynamics of the most influential institutional and organizational conditions as noted by scholars and practitioners,

and how and when they have influenced agency institutionalization of performance management reforms during this period. The purpose of this chapter is, thus, to study the influence of the most important conditions during the three primary phases of the modern history of performance management reform in the U.S. Government in order construct a contextual background and improve the framing and quality of the analysis and interpretation of findings from the present dissertation research in subsequent chapters.

To fully institutionalize GPRAMA in U.S. Government organizations, this study argues that federal agency[18] leaders have been required to accomplish tasks in four main categories. First, they needed to define and articulate clear policy goals and program objectives as part of a strategic planning process (Chun and Rainey 2005b; J. W. Lee, Rainey, and Chun 2009). Second, they had to develop capacity and analytic methods to perform valid, reliable, and objective internal and external *ex post* evidence-based evaluations and external reporting of achieved agency and program performance against goals suitable to inform budgeting decisions (U.S. Library of Congress, Congressional Research Service 2004; 2005). Third, to move beyond simple legal compliance in their response to the reform policy, agency leaders had to embed performance management practices and procedural knowledge into their line business units by establishing learning forums and routines to measure, report, and internally use program and organizational performance information to inform a range of program and management decisions (Moynihan and Kroll 2016). Finally, they needed to instantiate mechanisms for single- and double-loop learning[19] using the results of internal and external evaluations as well as the

[18] Unless otherwise specified, references to "agency" in this study imply federal departments and independent agencies, not merely subordinate elements of departments and bureaus.
[19] Argyris (1991) first proposed the concepts of single and double-loop learning. In short, he frames single-loop learning as problem solving, whereas double-loop learning entails critical

feedback of political leaders and appointees, oversight authorities, program stakeholders, program partners, and interested citizens to evaluate and continuously improve the effectiveness of their performance management systems (Bouckaert and Halligan 2008, 119-120). Decisions by federal agency leaders in these four categories have varied greatly between agencies during the modern era of performance management, resulting in wide variance in reform institutionalization in federal agencies today.

As discussed in chapter 1, recent evidence and analysis from GAO suggests institutionalization of GPRAMA may have reached its culmination, at least in the framework of current governance and agency structures and cultures. Further institutionalization of GPRAMA may demand changes in either or both of these, which are difficult to realize. Alternatively, as this chapter will show, a new tide of government reform may now be swelling that builds on the sedimentation of institutionalized GPRAMA practices and moves on to the use of new methods and means for encouraging government transparency, accountability, and performance.

The primary focus of reform legislation in this study is, of course, the Government Performance and Results Act Modernization Act of 2010 (GPRAMA), which is celebrating its 10-year anniversary even as this chapter is being written. But, GPRAMA was merely a significant refinement of the original Government Performance and Results Act of 1993 (GPRA)[20] which was ". . .the starting point and linchpin of the contemporary federal performance system" (Moynihan 2013, 7). GPRAMA statutory requirements reflect government-wide and stakeholder learning as of 2010 about the limits and challenges of

---

self-reflection on how individuals and organizations make decisions that lead to behavioral change (99-100).

[20] Government Performance and Results Act of 1993, Pub. L. No. 103-62, §1, 107 Stat. 285 (1993). https://www.congress.gov/103/statute/STATUTE-107/STATUTE-107-Pg285.pdf

performance management reforms. This learning was based on shared insights from single and integrated double-loop learning cycles during GPRA implementation (Bouckaert and Halligan 2008, 125, 400). The current state of GPRAMA institutionalization likewise reflects the collective learning and responses of legislators, staffs, political appointees, federal managers, and the public that transpired since it was signed in 2011 as well as all previous knowledge acquired after GPRA promulgation. Therefore, after 28 years since GPRA was promulgated, this study argues that observed variance in the embeddedness and forms of routine performance management practices (i.e. its institutionalization) between federal agencies is best explained as the product of multiple tacit or explicit compromises between external and organizational stakeholders to accept something less than uniform government-wide implementation. These compromises implicitly acknowledge the role of the system action and organizational conditions on the ability and willingness of federal agencies to implement GPRA and GPRAMA.

Barring new legal authorities, or more innovative management approaches to institutionalizing performance management reforms, what we have today may well represent the limit of what is possible for the institutionalization of GPRAMA, as a freestanding piece of legislation, in the context of the vast and complex bureaucracy of the U.S. Government. But, how did we get to where we are today, and what does the history of the performance management reform since 1993 suggest about the power of longstanding system action and organizational conditions on federal agency support for, or resistance to, institutionalizing government reforms generally, and GPRAMA in particular?

The 30-year history of GPRA, PART, and GPRAMA is the story of a sustained and successful public management reform movement that, in the final analysis, has survived repeated rounds of attack by philosophical opponents and changing institutional environments by its

negligible administrative costs and the face value of its almost universally-accepted and popular good government aspirations. These aspirations include improving federal budgeting; agency decision making, transparency, accountability, performance, and citizen satisfaction with quality products and services. Its promise, or at least potential, to reduce the cost of federal programs to taxpayers has been like catnip for conservatives of both parties in Congress. From a rhetorical standpoint, it is difficult to argue against the merit of these policy goals when framed in such an idealistic way, as proponents have been largely successful in doing.

What many early critics denounced as just another management fad that would ultimately die a similar death as the Planning, Programming, and Budgeting System (PPBS), Management by Objectives (MBO), and Zero-Based Budgeting (ZBB) has turned out to be a durable performance reform movement whose continual development was influenced by lessons learned from its many failures along the way. That the nation's political leadership at different times chose to use those failures as learning opportunities to get the next phase right, and not to throw the reform out for short-term political payoffs, testifies to both their sense of public service responsibility and to the underlying validity of the requirement to continuously improve the performance of an immense, highly-regulated, and path-dependent bureaucracy wherever and whenever it is feasible to do so. In addition, performance management reform was also seen by many as a strategy to help address fiscal and political pressures from rising federal deficits.

For some federal agencies, the story of their implementation of performance management reform is one of poor cultural fit, steep learning curves, competition with other policy goals, managerial conflicts of interest, and motivated behavior and satisficing. In other cases, agencies experienced resource limitations, setbacks during implementation, and skeptical workforce attitudes towards performance management. The result was often agency responses of simple

legal compliance and passive performance information use. For other agencies, it is one of adaptation; entrepreneurial, committed, and creative leaders; and real organizational transformation through double-loop learning, employee training, and investments in new evaluation capacity and analysis methods. These experiences led to agency strategies that deeply institutionalized performance management routines using performance information actively that were valid, widely accepted, standardized, embedded deeply across and down the organizational hierarchy, and used by managers for internal program decisions. Most agencies were situated somewhere between these two ideal types. Agencies with organizational identities that enabled them to measure the results of their programs had fewer problems adapting to GPRA and GPRAMA requirements than agencies whose program outcomes were difficult to define with precision and/or measure quantitatively. The history of the modern performance management reform era helps to reveal why this was so, and to make sense of the challenges and opportunities federal agencies face today as they continue to try to institutionalize GPRAMA and more recent evidence-based reforms.

As data sources for this chapter, I mined archival and secondary sources, including what is arguably the leading scholarship on performance management policy implementation. I also supplemented these with interviews of public management experts at NAPA who were directly involved with GPRA, PART, and GPRAMA development and implementation or scholarly research to characterize the mediating roles that the system action and organizational conditions in this study played. Following methodological recommendations of historical institutionalists Richard Neustadt and Ernest May (1986) to understand and analyze organizational history using timelines (231), this historical evaluation will follow approximate chronological order to provide insights into the complex patterns and effects of system action and organizational conditions that

contributed to the uneven and iterative nature of performance reform institutionalization. These historical insights will provide a foundation for comparing, contrasting, and interpreting the findings from empirical data about federal agencies collected for this study in chapters 4 and 5, and allow lessons from the past about organizational cultures and identities to suggest better ways for policy makers, federal managers, and other stakeholders to govern performance management reforms in the future.

The analysis unfolds over four successive periods spanning the Clinton, Bush, Obama, Trump, and early Biden administrations: 1) the origins of government-wide performance management reforms, 2) socializing, testing, and learning to measure during GPRA from 1993-2000, 3) experimenting with performance budgeting with PART from 2001-2009, and 4) strengthening institutional governance with GPRAMA from 2010 to the present. Because party control of the Presidency and Congress played a somewhat significant role in all of these periods, table 3.1 shows the history of political party control of the White House, Senate, and House of Representatives, by numbered Congress, since 1991:

Table 3.1. Political Party Control of Presidency, Senate, and the House Since 1991.

Years	Presidency	Incumbent	Congress	House	Senate
1991-93	R	George H.W. Bush	102	D	D
1993-95	D	William J. Clinton	103	D	D
1995-97	D	William J. Clinton	104	R	R
1997-99	D	William J. Clinton	105	R	R
1999-2001	D	William J. Clinton	106	R	R
2001-03	R	George W. Bush	107	R	D
2003-05	R	George W. Bush	108	R	R
2005-07	R	George W. Bush	109	R	R
2007-09	R	George W. Bush	110	D	D
2009-11	D	Barack H. Obama	111	D	D
2011-13	D	Barack H. Obama	112	R	D
2013-15	D	Barack H. Obama	113	R	D
2015-17	D	Barack H. Obama	114	R	R

2017-19	R	Donald J. Trump	115	R	R
2019-21	R	Donald J. Trump	116	D	R
2021-Now	D	Joseph R. Biden	117	D	D

Legend

R	Republican Control
D	Democratic Control

Source of data: https://www.spokesman.com/stories/2020/jun/25/control-house-and-senate-1900/

For the first period, 1993-2000, the study describes the intellectual and political origins of the modern performance management movement at the federal level of the U.S. Government and the public policy process that led to the promulgation of the Government Performance and Results Act of 1993, the antecedent legislation for GPRAMA. For the remaining periods, the analysis will examine evidence about the relative influence of the system action and organizational conditions in the study's inferential framework on GPRA, PART, and GPRAMA institutionalization through the eyes of stakeholders, political appointees, federal managers, and oversight authorities who were involved. The chapter concludes with a summary of the latest developments in GPRAMA governance and a forecast of the way ahead for performance management-related reforms.

The Origins of Government-Wide Performance Management Reforms

In its most recent incarnation, the primary reform movement in the U.S. Government since the early 1990s, which Light (1997) dubs "liberation management," has aspired to help federal managers lead their agencies to achieve higher levels of performance:

> . . .liberation management is unabashedly immodest in pursuit of higher performance. It is through the setting of customer service standards, the establishment of measurable goals, the assessment of actual outcomes that liberation management finds the confidence to free agencies from the burdens of central controls. Its notion is to pull government upward by raising the aspirations of agencies and employees, and by measuring their progress toward success (39).

Previous government reform movements had primarily valued economy, efficiency, integrity, political neutrality, and public participation in government operations. The management reform program of the Reagan Administration, Reform-88, had emphasized the prevention of fraud, waste, and abuse, cost savings, cash and credit management, and the integration of multiple accounting systems (President of the United States 1989, 1-12). Program evaluation had once been widely practiced in the federal government in the 1960s and 1970s, but the function was decimated by declining budgets during the 1980s (Weinstock 2003, 40). Whence, then, emerged the predominant policy value and agenda of improving and measuring government performance? Light (1997) offers an explanation from the standpoint of a political scientist by crediting beliefs and policy preferences of Thomas Jefferson; Defense Secretary McNamara's Program, Planning and Budgeting System (PPBS); President Nixon's cabinet reform proposals; President Ford's Management by Objectives (MBO); and President Carter's Zero-Based Budgeting (ZBB) rules (38-39). Newcomer and Brass (2021) see in MBO an inspiration for the rise of centralized goal-setting and performance measurement values during the GPRA era of the 1990s (51). NAPA credits PPBS and ZBB with setting the stage for the expansion of knowledge about program budgeting and program analysis (National Academy of Public Administration 1999, 8). The historical analyses of Light, Newcomer, Brass, and NAPA, while accurate, fail to fully satisfy because they do not portray the transfer and diffusion of dynamic societal values or the emergence of the philosophical paradigm shift that ushered in and propelled the NPM reform era. Two values of liberation management, competition and deregulation, have nevertheless inspired several scholars to investigate the social origins of performance management philosophies as evidence of a deeper political and social transformation in American society.

W. Richard Scott (2017) traces public-sector performance management reforms to the historical rise of the neo-liberal institutional logic beginning in the 1930s with Hayek and Austrian economists, later coming into full theoretical bloom through the writings of the University of Chicago school of economics and Milton Friedman, who emphasized cutting regulations and taxes while challenging traditional logics of professionalism (860-861). The late anthropologist David Graeber (2015) argued the adoption of performance management concepts like "vision," "quality," "stakeholder," "leadership," "excellence," "innovation," "strategic goals," and "best practices" in government was evidence of the diffusion of popular management values and techniques from corporations and financial institutions which had their roots in 1970s-era management fads that encouraged managers to "self-actualize" (21). By 1993, such private-sector management reform concepts began to enjoy acceptance among many thought leaders in the public sector after the publication of Osborne and Gaebler's (1992) "Reinventing Government" (Green and Hubbell 1996, 39-40). This popular and influential book encouraged, *inter alia*, the use of market-based incentives, competition between public agencies and between public agencies and the private sector in the provision of public services, public managerial entrepreneurship, and organizational decentralization as principles to establish "entrepreneurial government" (Green and Hubbell 1996, 41). As described below, it became the doctrinal North Star for performance management reforms during the Clinton Administration and continued to serve that purpose for subsequent variations of performance management reform.

Finally, several sources credit the influence of the positive experience several foreign governments had using performance management methods prior to and during the early 1990s in which they associated clear goals and measures with accurate cost accounting to improve program outcomes, especially in Australia, New Zealand, Canada, the United Kingdom, and

Sweden (National Academy of Public Administration 1999, 8-9). One noteworthy example from Great Britain, cited in testimony by Charles Bowsher, Comptroller General of the United States, before the Senate Governmental Affairs Committee, was the adoption of new management flexibilities for the British Employment Service to permit managers to realign funds between appropriation accounts and carry funds over to future years in exchange for increased accountability for meeting quantitative targets for job placement for unemployed people and accuracy in payments (U.S. General Accounting Office 1992a, 6-7). American state governments, in particular Florida, Minnesota, Oregon, and Texas, had also started making inroads in the use of performance management prior to GPRA (U.S. General Accounting Office 1992a, 15).

A management reform philosophy without any champion or mandating legislation is just a passing fad, or slogan, at best. Real institutionalization of reform in government requires legislation. Serious reform actors and coalitions, of course, are well aware of this and look for opportune moments to advance their reform agendas. While some scholars suggest the first legislative mandate for U.S. agencies to conduct performance management was the Chief Financial Officers Act of 1990 (Kamensky 2021), the first notable impetus for the institutionalization of government-wide performance management reforms in the U.S. Government began in 1993 with the Government Performance and Results Act (GPRA) of 1993.

The idea to legislate a performance management framework to improve managing for results at the federal government level did not originate in the United States. According to Scott Fosler, then President of NAPA, by the time GPRA was passed, many foreign governments, including Australia, New Zealand, and the United Kingdom, had already responded to economic pressures by instituting reforms of this nature, and the U.S. Government was trying to catch up

(Council for Excellence in Government and National Academy for Public Administration 1998). The U.S. learned from the mistakes and accomplishments of Australia and other Organization for Economic Cooperation and Development (OECD) countries in developing its managing for results philosophy (Breul 1994, 2). This suggests institutional isomorphism, in the form of international diffusion of performance management reform, helped to change the perspectives of U.S. legislators to better align with those of good government reformers in support of performance management institutionalization.

Local governments in the U.S. had also instituted and successfully used models of performance management systems for several years. The influence of their example at the federal level should not be discounted. As the Counsel to the Senate Governmental Affairs Committee from 1989-1997 under senators William V. Roth (R-DE) and Ted Stevens (R-AK), John Mercer claimed the mantle as the leader for the development of GPRA (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001, 63-65). Mercer had previously served as a councilmember and then as mayor for the city of Sunnyvale, California, whose advanced performance management and budgeting systems inspired him to recommend it to U.S. senators and congressional staffs as a model for the federal government (Mercer 2001, 5-6). His message resonated with conservatives in Congress, who threw their support behind GPRA (Lynn 1999, 3), and it emphasized the value of legislation that would foster performance budgeting (Radin 2000, 116-117). GPRA enjoyed bipartisan endorsement from the outset. According to Fosler, the legislation was co-sponsored in the Senate by Senator William Roth, a Republican, and John Glenn, a Democrat, and by members of both parties in the House (Council for Excellence in Government and National Academy for Public Administration 1998). He noted as well that after

the election of 1994, in which Republicans gained control of the House and Senate, Republicans reaffirmed their support. During the early days of GPRA implementation, Senator John Glenn summarized Congress' expectations for the law:

> Some people view GPRA as a tool for reinventing government. Others view it as a weapon for weeding out unneeded programs. I view it simply as a good government attempt to shift the attention from inputs and outputs, like budget dollars and [full-time equivalents], to outcomes, to the real results that Americans care about. We are talking about results like saving lives, improving literacy, providing quality health care for all, setting the specific objectives and ways of measuring whether we are accomplishing those things or not (U.S. Congress. Senate. Committees on Governmental Affairs and Appropriations 1997, 6).

For legislators who were comfortable with analyzing agency budgets and resource inputs to programs as part of the well-trod budgeting and appropriations processes, the notion of trying to measure outcomes of federal programs, as a first step to evaluating their cost-effectiveness, was probably the most revolutionary change that GPRA introduced into the history of U.S. Government reform. It raised public accountability expectations to a whole new level. One can see the ambitiousness of this expectation most clearly in the National Performance Review (NPR) of 1993, led by Vice President Al Gore. In making a strong case for the use of outcome-based performance measurement, the NPR benchmarked Sunnyvale's successes, and used them as a lever to open a new public discussion on introducing new managerial incentive structures and flexibilities into the federal public sector that would be familiar to private sector managers:

> Outcome-based management is new in the public sector. . .In each policy area, [Sunnyvale] defines sets of "goals," "community condition indicators," "objectives," and "performance indicators'. . .Sunnyvale measures performance to reward successful managers. If a program exceeds its objectives for quality and productivity, its manager can receive a bonus of up to 10 percent. This generates pressure for ever-higher productivity. The result: average annual productivity increases of four percent. From 1985 to 1990, the city's average cost of service dropped 20 percent, in inflation-adjusted dollars. According to a 1990 comparison, Sunnyvale used 35 to 45 percent fewer people to deliver more services than other cities of similar size and type. . . . (National Performance Review 1993, 76).

Compelling examples and ideas for good government like this may float around in the public consciousness and dialogue for years without legislative action to transform them into law or policy. How did performance management avoid this fate? The promulgation of GPRA in 1993 serves as a useful case study of the policy adoption process conceptualized by John W. Kingdon, which he acknowledges is an application of Cohen, March, and Olsen's (1972) "Garbage Can" model of organizational decision making for the public policy process (Kingdon 2003, 84-88).

Kingdon's "three streams" theory (problem, politics, and solution, in no particular order) offers a useful mental model to study the origins of GPRA during this window in a way that exposes the causes for the emergence of the performance management reform policy agenda and offers a starting point for studying the legal, oversight, and political conditions in the system action environment. John Mercer, whose profile fits quite comfortably with the theoretical descriptions of a "policy entrepreneur" by both Kingdon (2003, 179-182) and the Advocacy Coalition Framework (Jenkins-Smith et al. 2014), successfully advocated for GPRA by connecting the streams of problems and politics with a novel policy solution stream during a brief period in the early 1990s. The federal election of November 1992 opened a policy window, which Kingdon (2003) defines as "…an opportunity to push pet proposals or one's conception of problems…" (194), for serious consideration of performance-oriented government reform by both the Executive Branch and Congress.

GPRA was conceived as a means to address real problems. While there was no 9/11-style terrorist attack or Katrina-style natural disaster on American soil just before GPRA, the Senate Committee on Government Affairs GPRA Conference Report to accompany S.20 of June 16, 1993 cited several problems to justify the GPRA legislation. It cited public opinion polls showing widespread perceptions of rampant waste in federal government spending, inefficiency,

and ineffectiveness which the committee alleged had led to a crisis in public confidence in U.S. governmental institutions (U.S. Congress. Senate. Committee on Governmental Affairs 1993). Indeed, the same report attributed much of the blame to certain federal agencies which the committee believed lacked a results orientation (U.S. Congress. Senate. Committee on Governmental Affairs 1993). In the enacted GPRA legislation, Congress found that:

> (1) Waste and inefficiency in Federal programs undermine the confidence of the American people in the Government and reduces the Federal Government's ability to address adequately vital public needs
>
> (2) Federal managers are seriously disadvantaged in their efforts to improve program efficiency and effectiveness, because of insufficient articulation of program goals and inadequate information on program performance; and
>
> (3) Congressional policymaking, spending decisions and program oversight are seriously handicapped by insufficient attention to program performance and results.[21]

Vice President Albert Gore echoed these concerns by arguing there had been a ". . .a steady erosion and decline in the confidence that the American people have had in our self-government. . ." that bred distrust and cynicism of institutions, and indicated there were several studies that attributed declining public confidence to poor government performance and waste (Council for Excellence in Government and National Academy for Public Administration 1998). GPRA offered a potential solution to these problems. A core premise of the reform was that getting federal agencies to focus more on results would ultimately lead them to change their behavior in ways that would improve program performance. According to Fosler, improving public confidence in government by better serving the public was an important objective of GPRA (Council for Excellence in Government and National Academy for Public Administration 1998). The Senate committee on Governmental Affairs claimed GPRA could

> . . . mark a significant change in the way that managers, policymakers, and the American people think about what services the government should provide, and how well it does at

[21] Government Performance and Results Act of 1993.

> providing them. [GPRA] will provide the information necessary to strengthen program management, to make objective evaluations of program performance, and to set realistic, measurable goals for future performance--ensuring that the information is reliable will, of course, require attention by agencies, OMB, GAO, and Congress (Senate Committee on Governmental Affairs 1993, n.p.).

A change in the political stream opened the policy window. The election of William Clinton as President and Albert Gore as Vice President in November 1992 allowed the platform of the Democratic Party, developed in the summer of 1992, to implement its political agenda for a wide range of policy domains in the new Administration. One plank of the platform signaled a shift in the party's view of governmental accountability, seeking a new balance between the extremes the high public distrust of government during the Reagan Administrations and the public administration heyday of the New Deal and Great Society eras, when federal programs had high autonomy but minimal oversight:

> For twelve years, the Republicans have expected too little of our public institutions and placed too little faith in our people. We offer a new social contract based neither on callous, do-nothing Republican neglect, nor on an outdated faith in programs as the solution to every problem. We favor a third way beyond the old approaches—to put government back on the side of citizens who play by the rules (University of California at Santa Barbara. The American Presidency Project. n.d.).

That "third way" was later clarified at a rhetorical level: "We vow to make government more decentralized, more flexible, and more accountable—to reform public institutions and replace public officials who aren't leading with ones who will." (University of California at Santa Barbara. The American Presidency Project. n.d.). The history of the NPR from March to September 1993, and the "Reinventing Government" campaign associated with David Osborne and led by Vice President Gore, captured the spirit of the Democratic platform's "third way." It supports an interpretation that the new Administration found in performance management reform a politically-advantageous mechanism to reconcile the competing goals in the Democratic party platform of decentralizing government and increasing flexibility while sustaining an acceptable

level and form of accountability. The final NPR report to the President stated its goal was ". . .moving from red tape to results to create a government that works better and costs less" and that it aimed to ". . .transform the habits, culture, and performance of all federal organizations" (National Performance Review 1993).

Immediately prior to the GPRA era, a GAO evaluation in 1992 found that federal agency performance management systems (PMS) lacked an orientation on program results (or outcomes): while two thirds of agencies studied by the GAO had a strategic plan with goals and objectives, and three quarters collected data to assess performance at the program level, performance measures were used almost exclusively internally and seldom reported to OMB or Congress, and were not suitable for informing agency-level decisions on policy or strategy (U.S. General Accounting Office 1992, 1-2, 9). In addition, three indicators of a PMS were generally absent in federal organizations: a strategic plan with goals that could be measured, an organization tasked with collecting performance measures, and the production of consolidated reports (U.S. General Accounting Office 1992, 10). To imbue federal agency cultures with a results-focused orientation, new tools and techniques would be required. The NPR identified performance measurement as an important one, along with quality management and training, that would enable the government to reduce bloated legacy "management control structures" employing 700,000 federal employees and costing 35 billion dollars in salary and benefits, permitting the elimination of 252,000 positions, about 12 percent of the non-postal workforce (National Performance Review 1993, 13-14). Radin (2000) observed that the Clinton administration supported GPRA's goal of improving agency performance and saw Congress' GPRA as "complementary" to his NPR (116). More specifically, officials believed GPRA implementation would promote the NPR objective of encouraging agencies to release their

performance information (Breul and Kamensky 2008, 1014). According to Fosler, because GPRA aligned well with NPR principles, Vice President Gore urged President Clinton to support and sign the GPRA bill. (Council for Excellence in Government and National Academy for Public Administration 1998). The coalition of GPRA advocates extended beyond the government proper, but only modestly, and mostly inside the Washington Beltway. The American Society of Public Administration and NAPA also backed GPRA (U.S. Congress. Senate. Committee on Governmental Affairs 1993, 8-9). Fosler noted NAPA also sought to assist 26 federal agencies with implementing GPRA by establishing a performance consortium (Council for Excellence in Government and National Academy for Public Administration 1998).

The Clinton Administration's new reform agenda, ironically, favored a kind of Reaganesque "trust but verify" philosophy towards federal programs and their managers. The Administration would demonstrate greater trust towards agencies and managers by eliminating organizational layers, oversight officials, and procedural red tape associated with traditional compliance oversight of programs. It would continue to verify performance through mandated mechanisms of transparency such as agency strategic plans and performance reporting as acceptable means to continue to hold bureaucracies accountable, or at least appear to do so. In principal-agent theory, principals attempt to minimize the costs associated with principal-agent conflicts of interest in an organization by setting rules and instituting procedures to govern agent incentives (Thaler 2015, 105). Because of its focus on the use of incentives to control managers' decisions, principal-agent theory offers a suitable lens for interpreting the doctrinal foundations of GPRA which similarly emphasized new governance systems to shape manager incentives.

Performance measures are a useful tool for principals to offset the advantages informational asymmetries give to agents who control essential information about the programs

they manage (Moynihan et al. 2011, i142). Using techniques of performance measurement and reporting, GPRA sought to establish a new relationship between principals (here, U.S. political leaders and appointees) and agents (here, federal managers) that would reduce the costs of administrative oversight of federal programs to principals and empower federal managers with greater autonomy and authority to drive higher performance in their programs while maintaining an acceptable level of accountability using less costly and intrusive means. But what exactly is an "acceptable" level of accountability, and how did the new accountability regime differ from the old one? NPM did not mean less accountability for federal agencies, just a different kind.

Romzek and Dubnick (1987), in their widely cited study of the causes of NASA's 1986 Challenger disaster, propose a taxonomy of four forms of accountability which vary based on two factors: the degree of control over agency actions (high or low), and the source of control exerted on bureaucracies (external or internal). When control is imposed externally through laws, contracts, or political influence, then legislators use legal accountability to maintain a high degree of control, and political accountability, in the form of responsiveness to citizens, when a lesser degree of control is desired. When agencies are accountable internally, then bureaucratic accountability using close supervision within the agency's leadership hierarchy is used for high control, and professional accountability among technical experts using their knowledge of programs, with some influence by outside professional organizations, for less control. Both forms of accountability with a low level of control are associated with high agency discretion (Romzek and Dubnick 1987, 228-229). Furthermore, Romzek and Dubnick (1987) hypothesize that ". . .if the environmental changes are drastic enough, they may trigger a different type of accountability system, one which attempts to reflect those new institutional conditions" (230).

Under GPRA, the new legislation and the changing political agenda in the Clinton Administration drove a shift in accountability arrangements for federal agencies and managers, often dramatically, offering empirical support for Romzek and Dubnik's claim. From the 1960s through the 1980s, the primary form of accountability for federal agencies was the "watchful eye" approach, as Light (1997) described it, to prevent and detect fraud, waste and abuse in bureaucracies. This accountability model subjected federal agencies to extensive external oversight and control, leaving their managers responsible only for compliance with anti-fraud, waste and abuse statutes but little discretion (Gormley and Balla 2008, 13). While there were doubtlessly elements of political accountability involved with this, and "watchful eye" measures were not suspended during the Clinton years, watchful eye focused on legal accountability via external control networks and actors. Watchful eye oversight fosters a compliance culture among federal managers, robs them of creative energy, and turns them into passive instruments of the state.

With the dawn of GPRA, perhaps in part due to the increasing complexity of federal programs (Kettl 2016, 189), political leaders began to rely less on legal and political forms of external accountability in favor of internally-focused bureaucratic and professional accountability and the use of goal setting, capacity building, incentives, technological solutions, and de-layering of hierarchical organizations to manage an expansion of bureaucratic discretion. The Reinventing Government reform philosophy depended on a substantial expansion of bureaucratic discretion to improve organizational performance. For example, during the development of the GPRA legislation in 1992, Comptroller General Charles Bowsher argued that

> Managers must be given the incentives and tools to create a results-oriented work environment. Without incentives and tools to manage for results, managers have little control over their operations. Consequently, they focus more on complying with the accountability controls placed on them (which often have multiple objectives) than on

> achieving their agencies' overall missions. These controls affect their hiring, promoting, and rewarding; create a rule-intensive procurement and contracting maze; and rigidly regulate funds." (U.S. General Accounting Office 1992a, 29-30).

Along with transparency through periodic reporting to Congress, these tools would become the accountability cornerstones of liberation management (Light 1997, 39, 52-54). Instead of regarding government as "the problem," as Reagan announced at his first inauguration,[22] the Clinton Administration portrayed government, and federal workers, as a fundamentally trustworthy source for policy solutions. Indeed, as Vice President Gore stated, GPRA was designed with the intention of creating incentive structures for those very same public managers (Council for Excellence in Government and National Academy for Public Administration 1998).

On the other hand, the author has found no historical evidence the reformers inquired if federal managers would even desire to exchange longstanding arrangements of compliance accountability and minimal discretion for performance accountability and significant discretion, as managers in the private sector might welcome. It appears GPRA and GPRAMA reformers either took the support of federal managers for the reform for granted, or, if they considered their views at all, acted on an assumption that federal managers tasked to implement GPRAMA would consider the grants of additional managerial discretion to be well worth the additional costs of managing and reporting performance and the potential professional hazards of being held accountable for "poor" performance outcomes, possibly beyond their control. Why would this occur, given the new NPM-related espoused values of trust in and delegation to public officials in the 1990s, and why might federal managers impede policy institutionalization?

Behn (2001) explains that public managers have logical reasons to avoid participating in what he terms "responsibility compacts" (142). He argues such compacts may leave public

[22] https://www.reaganfoundation.org/ronald-reagan/reagan-quotes-speeches/inaugural-address-1/

leaders and managers vulnerable to receiving concentrated blame for the failures of the performance consortium based on the allegations of other officials in the compact who, due to their response to being in a prisoner's dilemma with other compact members, opt to defect just before or after a public disclosure of performance failure (144-149). According to Tummers, Steijn, and Bekkers (2012), many additional factors affect the willingness of public managers to implement policies; two important ones are the level of discretion and manager views of the meaningfulness of the policies they are tasked to implement for society, their clients, or themselves (719). They found statistically significant positive associations for discretion[23] but negative associations between public Dutch public employee perceptions of all three forms of policy meaninglessness and their willingness to implement a controversial new national healthcare reform law (727).

Based on this theory and research, it is therefore arguable the NPM reformers of the 1990s should have considered that some U.S. Government managers might sense personal or professional risks from participating in a risky policy implementation project when previous waves of reform experiments had failed. Alternatively, they might have considered federal leaders and managers might perceive varying levels of policy meaninglessness in GPRAMA based on the mission frame of reference of their organization or program, and that in turn their views could detract from their willingness to implement it. At its root, such an oversight may reflect a lack of understanding that federal managers possess significant agency in the policy

[23] The coefficient for discretion was significant only at the .05 level and had a small coefficient, suggesting the net effect of discretion is only slight positive and not especially important. The authors note in the theoretical discussion that discretion can swing both ways: a high level of discretion in the context of low policy support can lead to very low willingness to implement policy.

implementation process, and so should be included in the network of stakeholders which formulate reform policies.

This new approach to performance accountability was, nevertheless, enshrined in the form of new requirements and opportunities for federal managers in the enacted GPRA statute. A multi-year timeline for implementing GPRA requirements was established. Agency heads were required to produce and deliver five-year strategic plans for their programs to OMB and Congress by September, 1997 (section 3), and to submit annual performance plans with "objective, quantifiable, and measurable" performance goals[24] and output, outcome, and service level performance indicators for budgeted program activities to OMB starting fiscal year 1999 (section 4). Agencies also were required to start submitting prior year annual program performance reports to the President and Congress beginning in March, 2000 that would compare achieved program performance with performance goals. Except for the reality that federal agencies did not have required cost accounting systems to provide cost data for programs, some reformers thought GPRA might also have included a requirement for agencies to report costs per unit of result, service, or output, all measures of program efficiency (Mercer 2001, 10). The absence of cost accounting systems is therefore evidence that organizational conditions impacted on the reform legislation even during policy development, since it restricted options for new requirements in the legislation for program efficiency measures.

For individual federal managers, GPRA section 5 offered the legal foundation and opportunity for a more flexible approach to managerial accountability with a grant of greater discretion. By affording federal managers the opportunity to request waivers for policy

[24] If performance goals were too difficult to measure, agency heads could petition OMB to authorize them to use of an alternative form of measurement involving descriptive statements.

constraints on resource management, it accorded with the later analysis of Moynihan (2008) on the need for greater managerial authority to enable effective performance: ". . .performance management doctrine. . .emphasizes that efficiency gains through process improvement are inherently limited by constraints on managerial authority. Likewise, results-based accountability will not occur when managers lack authority over processes and, ultimately, outputs" (47-48). Accordingly, GPRA section 5 created a waiver process that would potentially grant federal program managers these special authorities, on a case-by-case basis, starting in fiscal year 1999, by allowing them to request of OMB "waivers of administrative procedural requirements and controls. . .in return for specific individual and organizational accountability to achieve a performance goal."[25]

OMB-approved GPRA waivers would grant federal managers some flexibilities from traditional organizational controls on staffing levels, personnel compensation, and funding transfers between budget lines, and permit them to make dynamic and resource-optimizing decisions like managers in the private sector are often able to do. Theoretically, this provision would "liberate" federal managers from bureaucratic constraints that prevented them from managing for results and improving performance. It was arguably a necessary condition to institutionalize liberation management reform in the U.S. Government because, without such flexibilities, federal managers would lack the discretion to optimally allocate scarce resources in a way that optimizes agency outputs and outcomes, but by itself it was not sufficient. Extending the values of New Public Management that advocated public management practices more aligned to those used in the private sector and under competitive conditions (Box et al. 2001, 612) into federal agencies, such management flexibilities could, if actually granted, theoretically foster

[25] Government Performance and Results Act of 1993.

entrepreneurship and risk-taking in public sector program management (Light 1997 as cited by Frederickson and Frederickson 2006, 37). The Senate Committee on Governmental Affairs cited precedents for use of liberal management flexibilities:

> The Committee encourages agencies to be creative and entrepreneurial in developing and applying managerial flexibility waivers. The successful experience of other national governments and certain state and local governments, in providing much greater authority to managers and staff in administering and implementing programs, suggests that substantial improvements in performance can result. *A limited or constrained approach to waivers is unlikely to lead to much improvement in performance* (Senate Committee on Governmental Affairs 1993, italics added).

New flexibilities for line and program managers, of course, can pose a threat to the authorities and prerogatives of managers in associated management functions such as finance and human resources who are responsible for controlling these functions under a range of laws and policies unrelated to GPRA. Perceiving greater autonomy for line managers as a challenge to their bureaucratic responsibilities and standing, financial and human resources officers in government agencies may oppose relaxing managerial controls (Moynihan 2008, 71). The National Academy of Public Administration (1999) frames the institutional conflict borne of parallel legislative mandates with commendable clarity:

> The requirements in law, except for the Results Act, focus on particular administrative disciplines, such as finance or information technology. Those offices sometimes build 'concrete silos' around their interpretation of the requirements of a particular piece of legislation and their role in its implementation, to the detriment of general management, achieving program results and an integrated agency approach to the implementation of all of the laws. (6)

As I will show shortly, such internal regulatory conflicts emerged during the early years of GPRA implementation, have never been fully reconciled since then, and have impeded GPRAMA institutionalization. The absence of legislative changes to deconflict the responsibilities of these management functions in a way that encourages agency realignment to a stronger results orientation is a testament to the power of institutional path dependency and the

sedimentation of institutional norms supporting compliance-focused oversight through these traditional control channels.

Besides a more trusting approach to accountability and new administrative flexibilities for managers, an additional GPRA aspiration was to foster reform experimentation and learning to support future implementation by mandating agency-level pilot projects across a range of functions under OMB oversight. This GPRA provision represents the institutionalization of double loop learning for the GPRA performance management regime. Ten performance management pilots were required in fiscal years 1994-1996, five for managerial accountability and flexibility in fiscal years 1995-1996, and five for performance budgeting in fiscal years 1998-1999 (Pub. L. 103-62, section 6). The first set were designed to "[test] and [demonstrate] whether the specifications and structure for the annual performance plan and program performance report work as intended," the second aimed to ". . .assess the effect of giving managers and staff greater latitude in administering and managing programs" and the third would seek to ". . .examine the practicability of determining and presenting the changes in performance levels that result from different funding levels" (Raines 1997, n.p.).

Summarizing the inception phase of GPRA, we can conclude from "top-down" policy implementation theory offered by Sabatier (1986, 23) that the policy launch of GPRA, while imperfect and incomplete in respects, met most theoretical requirements to be considered a success. First, it had arguably clear and consistent policy objectives with legal definitions for essential concepts. Next, it gave implementing officials, if they chose to make use of them, potentially powerful policy levers and incentives to use performance management to improve federal program performance based on the theoretical claim, then as yet untested at the federal level, that measuring and making transparent agency performance would ultimately improve it.

Finally, it mandated an implementation process that stepped away from external compliance-focused controls and emphasized greater internal bureaucratic and professional accountability by federal managers and agencies that depended on goal setting, performance reporting, and the use of performance information instead of heavy-handed compliance demanded by an army of oversight officials.

A comprehensive and balanced appraisal of the history of GPRA and GPRAMA implementation beyond its inception, however, requires a synthesis of "top down" and "bottom up" theories of policy implementation. Advocating the bottom-up view, May (1993) offers a flexible perspective on the principal-agent relationship in which government mandates may go beyond simply monitoring and enforcing agent compliance with goals and include ". . .various inducements, system changes, and capacity building features" and proposes that mandates are forms of "negotiated compliance" (as cited by Hupe and Hill 2016, 113). The bottom-up view assumes a higher degree of agency by public managers is possible than the top-down perspective does. Therefore, incorporating a bottom-up perspective permits the study to account for the effects of unique organizational conditions and managerial discretion to explain agency-level implementation decisions. Federal agencies and federal managers played a significantly more meaningful role during the implementation of the statutes than they did during their development.

Socializing, Testing and Learning to Measure During GPRA from 1993-2000

As noted, under GPRA agencies were required to produce five-year strategic plans, to be updated every three years, by September 30, 1997, annual performance plans by October, 1998 (the start of fiscal year 1999), and annual performance reports by March 31, 2000 and then again by the same date annually after that (Pub. L. 103-62, 107 Stat.285). In addition, GPRA required

a federal government performance plan by FY 1999. With the exception of the federal government performance plan, these deadlines served as milestones for agencies to institutionalize measures, evaluation capacities, and measurement and reporting processes in time to meet those statutory reporting dates far in the future.[26] In addition to levying new reporting requirements, GPRA established three sets of pilot programs from fiscal years (FY) 1994-1999 intended to foster organizational learning of capacity needs and procedures for writing performance plans and reports, applying performance budgeting methods, and waiving administrative controls on managers (U.S. General Accounting Office 1997, 4).

During the legislated phase-in period for GPRA, Vice President Gore sustained his support for the Reinventing Government movement (J. Kamensky and Chenok 2021). Federal employees implementing GPRA trusted President Clinton's internal messaging of commitment to the reform based on his previous commitments to Total Quality Management while Arkansas governor (Kamensky 2018, 307). Most legislators supported GPRA in the 1990s, though Republicans tended to support the "results" feature of the law more than the "performance" aspect (Radin 2000), and those who did wish to receive performance information were interested in using it primarily as a means to justify legislation to reduce duplication, fragmentation, and waste across federal programs (e.g. Thompson 2001, 57). Cooperation between the Legislative and Executive branches fostered a conducive political environment for federal agencies to develop capacities and routines to achieve a level of implementation that would at least ensure compliance with the new law. According to John Kamensky (2021), despite the so-called "Republican Revolution" in November, 1994 when the Republicans took control of the House of

[26] According to Jonathan Breul, OMB published one or two annual federal government performance plans in the early 1990s but discontinued it for unknown reasons (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 34).

Representatives for the first time since 1952, bipartisan support for GPRA was sustained, in part, by a collegial relationship between Vice President Gore and the new Speaker of the House, Newt Gingrich, who had previously served together in a science committee (interview with author 25 May 2021).[27]

Procedures for consultations between Congress and the Executive Branch on GPRA implementation took some time to formalize. From 1993 to 1997, prior to submission of the first agency strategic plans, consultations were limited to informational briefings, pre-consultations or preliminary consultations, and fraught with disagreements over expectations for the agenda for consultations, who the main participants should be, the number and scope of consultations, and requirements for delivery of preparatory materials (U.S. General Accounting Office 1997c). In 1998, House Majority Leader Dick Armey praised Congress' relationship with OMB and the performance of GAO on GPRA implementation matters (Council for Excellence in Government and National Academy for Public Administration 1998). However, traditional mechanisms of congressional oversight, which typically occur across a number of separate subcommittees which do not coordinate with each other to evaluate progress and performance by individual federal agencies, made it more difficult to enforce accountability. Vice President Gore noted agencies were under ". . .significant pressure to wall-off programs into specific budget accounts…or specific congressional subcommittee jurisdictions" (Council for Excellence in Government and National Academy for Public Administration 1998). Majority Leader Armey agreed this was a challenge: "Congress, in my estimation. . .comes nowhere near fulfilling its oversight

[27] It is probable this committee was the Congressional Clearinghouse for the Future. Heilemann (1995) notes Gore and Gingrich, while junior congressmen, were both members, shared a common interest in long-range scientific and societal issues, and developed a sense of mutual respect and a personal relationship (n.p). Thus, out of a common interest in science sprung a lasting relationship that would later help to protect the bipartisan consensus supporting GPRA.

responsibilities" (Council for Excellence in Government and National Academy for Public Administration 1998).

To overcome the problem of fragmentation in congressional oversight of GPRA,[28] Kamensky (2021) recalled that Speaker Gingrich, interested in holding federal agencies accountable for their strategic plans, agreed to establish new congressional task forces to integrate multiple authorization and appropriations committees overseeing agency development of these plans (interview with author 25 May 2021). The early history of GPRA, therefore, provides evidence of actionable feedback to Congress regarding GPRA implementation challenges, which it acted on, based on experience and organizational learning, as postulated in the inferential framework.

As the primary oversight entity for GPRA in the Executive Branch, OMB took the lead to direct and coordinate GPRA implementation across the interagency community. It led an educational effort starting in 1994 to socialize performance measurement and the use of strategic planning, produced a primer on performance measurement that included examples of different measures, conducted periodic reviews of program performance to generate more performance information that could be considered during reviews of FY 1997 agency budget requests, and led two reviews of GPRA implementation in 1996 (Raines 1997). Federal agencies were given a generous five-year window (six if annual performance reports are considered) to develop a full understanding of the reform and to start putting internal capacity and business processes in place to comply with the law. During an interview with the author, one former federal senior executive who was instrumental in crafting the original draft language for GPRA stated that the

[28] For example, Gore noted that EPA programs were overseen by 60 congressional committees and subcommittees (Council for Excellence in Government and National Academy for Public Administration 1998).

draft, written prior to the 1992 election, called for allowing agencies five years (1993-1998) to get started in part to shield the reform from potential political manipulation in case of a second Bush Administration.

A GAO evaluation in 1997 found mixed progress in GPRA implementation that pointed to possible impacts from several of the organizational conditions in the inferential model. It was already clear how steep the learning curve for agencies was. Despite scattered vignettes of improved performance in some programs,[29] many federal agencies had immature performance measurement systems and faced challenges with articulating their accomplishments, developing results-oriented performance measures, and using performance information to make decisions about programs (U.S. General Accounting Office 1997, 40). The first few years of GPRA revealed the widespread unfamiliarity of federal managers with the concepts and methods of strategic planning and how it is linked to organizational or program performance measurement, confusion over how to define performance goals and measurable indicators, uncertainty over how to measure agency outcome performance in the context of confounding external influences, and a common preference among federal managers to measure means or processes instead of achievement of program goals based on a concern about their lack of control over program outcomes (Raines 1997; U.S. General Accounting Office 1997b).

[29] Actual performance improvement was not a primary expectation of agencies during the phase-in period, but the promise of improved performance was a motivator for many stakeholders, so any early successes were significant to sustaining reform momentum and thus worthy of mention by GAO. Two impressive examples: the Coast Guard conducted a mission analysis that led it to emphasize education over regulation, a re-focusing that it credited for a decline in fatality rates in the towing industry from 91 per 100,000 employees to only 27 in just five years. The Defense Logistics Agency claimed dramatic improvements in inventory management through operational reengineering, reducing time to deliver medical supplies to customers from 30 days in 1993 to 24 hours by 1997 (U.S. General Accounting Office 1997, 44-45).

In analyzing the quality of the first round of agency performance reports submitted in 2000, Jerry Ellig at the Mercatus Center at George Mason University found many agencies presented a combination of activity and results goals, but it was unclear to him ". . . whether activity goals are offered out of desperation to find something measurable, or because the report's authors do not understand the difference between activities and results" (Ellig 2000, 7). Ellig (2000) credited the Department of Transportation with defining mostly results-based performance measures (e.g. fatality and injury rates, highway pavement conditions, energy efficiency, etc), but found that

> Many agencies measure virtually nothing but activities, including the Small Business Administration, Department of the Interior, Environmental Protection Agency, and Department of Health and Human Services. NASA focuses so much on internal project milestones that its measures may not be of use to anyone outside the agency. . . all but a few indicators at the Department of Justice track internal activities, such as prosecutions, training, investigations, sales of seized property, prisoners of victims assisted, and reinvention labs sponsored. (7-8)

An additional challenge cited by many federal managers in the early years of GPRA, which GAO argued could be partially addressed by increasing investments in program evaluation and other resources for performance analysis, was collecting performance data, especially when it was collected by non-federal entities, the accuracy and quality of performance data, and the timeliness of acquisition of data (U.S. General Accounting Office 1997b, 15, 30). Over the period, however, agency cultures evolved towards a results orientation through experiential learning and familiarity with GPRA philosophy. GAO engagements with federal manager focus groups revealed managers had begun to focus on performance measurement (especially on outcomes versus outputs and inputs) as well as program evaluation (U.S. General Accounting Office 2004, 12).

Federal career manager biases and conflicting incentives slowed GPRA implementation from the beginning. Long accustomed to compliance-based accountability, many federal managers were skeptical about the very notion of performance-based accountability. The Director of OMB explained: ". . .agency staff recurringly voiced concerns over being held accountable for achieving program goals where the actions of other parties will have much greater consequence" (Raines 1997, n.p.). Federal managers also debated with themselves whether or not to request additional resources to measure performance; asking for more could expose them to criticism that they had neglected to use performance information in the past, or were admitting they were using unsuitable measures (Raines 1997).

Federal managers looked to their senior leader political appointees for unambiguous signs of commitment to GPRA implementation strategies that would transcend mere legal compliance before deciding whether to pursue difficult changes in organizational culture or decision models; by 2000, they had not yet found them (Radin 2000, 125). GAO analysis shows this was not entirely the case, but hardly offered a ringing endorsement of senior leadership commitment, either: its surveys of federal managers found 53 percent of federal managers perceived a strong commitment from top agency leaders to achieving results in 2000, down from 57 percent in 1997 (U.S. General Accounting Office 2000b, 2). This evidence suggests many managers distrusted senior agency leaders' commitment to GPRA at that time, and illustrate Kettl's view that some federal managers see performance management as a kind of "gotcha game" because there is no way to win it (Kettl 2016, 155). To "win" in a performance management regime, public managers must believe that their senior leaders are committed enough to it to accept some risks and by protecting managers and their programs from unjustified criticism and budget cuts.

The pilot projects were a primary focal point for GPRA implementation during the phase-in period and their story deserves special attention because it reveals much about the effects of several organizational and institutional conditions. The urgent need to rectify the acknowledged insufficiency of agency capacity for evaluation and immature performance measurement systems was reflected in agency responses to the pilots. The GAO reported that 77 federal organizations participated in the first set of pilots on performance planning and reporting during fiscal years (FY) 1994-1996, more than GPRA required (U.S. General Accounting Office 1997, 40). Agency participation in these pilots included all 14 Cabinet departments plus 14 independent agencies, and about a quarter of the federal civilian workforce supported them (Raines 1997). OMB facilitated interagency co-production of knowledge and guidance on performance planning and reporting procedures for the first set of pilot projects using an OMB-led interagency task group that collaborated to write a series of guidance memoranda (Raines 1997). Though participation was strong and learning venues were institutionalized for these pilots, OMB found significant variance between agencies in terms of their ability to adapt to the new GPRA reporting requirements:

> Some agencies, often either because of their long-standing use of planning and measurement, or the relative facility in measuring their programs, may appear significantly advanced compared to others. At the beginning of implementation, it would be unrealistic to expect that there will be a uniformly high level of quality of agency plans and reports across the government. With time, and with the ability to compare and adapt successful approaches to performance management and measurement, the overall quality of agency plans and reports should improve significantly (Raines 1997, n.p.).

For the performance budgeting pilots, GAO noted the immaturity of agency analytic capacity and methods to associate program performance with varying levels of funding in most agencies (U.S. General Accounting Office 1997, 42). Congress and the OMB were postured equally as poorly to support performance budgeting. Radin (2000) noted appropriations committees in Congress were limited in their ability to make use of performance information for budgeting

reviews due to their relationships with their sub-committees and inadequate staff and time; that did not deter Republican leaders in Congress, however, from putting pressure on a minimally-staffed OMB management team to develop performance budgeting pilots (122-124).

The GPRA goal of five pilot projects to test additional management flexibilities foundered quickly on legislative and oversight constraints. This history reveals the often-invisible, yet very real, limitations of the institutional environment of the US Government which impose severe limits on the ability of agencies to empower federal managers to achieve higher organizational or program performance. OMB efforts to develop interagency procedures for review of management waiver requests exposed the effects of institutional and organizational conditions on the incentives of federal managers to implement this portion of GPRA. According to the GAO, management flexibility pilots from FY 1995-1996 ". . . did not work as intended" because OMB did not endorse any of the 61 administrative waiver proposals from 8 federal organizations (U.S. General Accounting Office 1997a, 3, 7). GAO assessed this was due in part to the 1994 Federal Workforce Restructuring Act that effectively prevented OMB from approving waivers requesting changes in personnel ceilings, and to the discovery that some agencies found they could leverage new NPR reinvention labs to improve program performance through business process reengineering, thus obviating the need for administrative waivers (U.S. General Accounting Office 1997, 41-42).

Federal law also severely limited what kinds of administrative waivers could be requested, and from whom. For example, statutory requirements could not be waived, nor could regulations governed by the Administrative Procedures Act without public review and comment as well as approval from the four federal agencies that administered them: the Department of the Treasury, the General Services Administration, OPM, and OMB (Raines 1997). OMB deemed

that all received waiver requests substantive enough to materially affect the activities of its associated pilot project would require a change in statute, therefore none were approved (Raines 1997). Oversight rules further limited the opportunity for agencies to pursue management flexibility pilots. OMB Director Raines (1997) reported that OMB and the other three regulatory agencies published a "metes and bounds for waivers" memorandum to all federal agencies that put them on notice that they would not entertain requests for waivers to certain administrative procedures, thereby potentially deterring entire classes of waiver nominations.

A final important reason for the lack of approved administrative waivers exposed the impact of organizational conditions: "when some agencies considered submitting pilot project nominations, they concluded that the requirements that most affected managerial flexibility were *self-imposed*, and not those originating in the central management agencies" (Raines 1997, emphasis added). In other words, the level of commitment of senior agency leaders towards performance management and agency cultures, not institutional legal or oversight restrictions, explained many cases of managers lacking the flexibilities they needed to better manage their funding or their personnel. Agencies had the authorities and resources their managers needed, but were tacitly acknowledging to OMB their inability or unwillingness to delegate them to managers. In the end, OMB authorized too few administrative waivers to justify a pilot project to test their impact on managerial effectiveness (Raines 1997).[30] Federal program managers could easily have concluded that they stood a better chance of gaining management flexibility by

[30] GAO criticized OMB for not consistently providing feedback, specific concerns, or clear instructions to agencies submitted waiver requests ". . . especially after it received those waiver proposals and concluded they did not meet GPRA requirements;" for its part, OMB officials believed agency waiver proposals ". . . were generally limited to seeking waivers from minor annoyances rather than significant barriers to improved performance" (U.S. General Accounting Office 1997a, 12).

advocating for it with their own agency leaders as by appealing to OMB and the interagency, or no chance at all.

By 2001, OMB Deputy Director for Management Sean O'Keefe nevertheless testified to Congress that departments and agencies were complying with GPRA and reporting their strategic plans, annual performance plans, and performance reports as required (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001). OMB complied with GPRA by publishing the first *Government-Wide Performance Plan* as part of the fiscal year 1999 budget (Executive Office of the President of the United States 1998). Unlike agency plans, the government-wide plan was not an expression of strategic goals for the American government, but a bottom-up administrative compilation of chapters of the FY1999 federal budget that loosely associated agency programs having similar budget functions and described program objectives of individual agency performance plans. The GAO criticized the report, noting its ". . . framework should ultimately allow for a cohesive presentation of governmentwide performance, but the specific contents of the initial plan did not always deliver an integrated, consistent, and results-oriented picture of fiscal year 1999 federal government performance goals" (U.S. General Accounting Office 1998, 2). GPRA implementation achieved a key milestone in the spring of 2000 when, for the first time, all CFO Act agencies completed the first full cycle of annual performance planning, goal setting, and performance reporting.

After the period of socializing and testing GPRA from 1993-1999, congressional leaders and Executive Branch oversight officials evaluated implementation progress in 2000 and 2001 based on the results of the first two annualized cycles of submission of agency performance plans and performance reports in 1999-2000. Their assessments indicated real progress had been

made, but legislative intentions for the reform were far from fully realized. For example, following initial congressional concerns regarding agency compliance with GPRA content requirements for their first draft strategic plans, GAO and congressional review teams determined that final agency strategic plans delivered in 1997 uniformly met the minimum legal requirements for compliance (U.S. Congress. Senate. Committee on Appropriations. 2000, 3).[31]

Annual GAO reviews of agency performance plans in fiscal years 1999 and 2000 also showed "moderate" improvement, but cited inconsistent focus on management challenges and risks to programs; insufficient cross-program coordination; and inadequate association of results with strategies, human capital, and management resources (U.S. Congress. Senate. Committee on Appropriations. 2000, 4). After reviewing agencies' fiscal year 1999 and 2000 performance plans, the GAO noted agencies provided only limited confidence in the credibility, reliability, and availability of their performance information due to 1) the lack of reported procedures agencies would use to validate and verify their information or address data limitations, 2) aspects of program design that delegated responsibility for performance measurement to state and local levels, and 3) the absence of agreed standards and definitions for concepts used in program evaluations (U.S. General Accounting Office 2000b, 5-6, 8-9).

Christopher Mihm expressed a more fundamental concern that agencies had shown they had difficulty "understanding how an agency's program and strategies contribute to results. . . this limitation has enormous implications for performance reporting " (Mihm 1999, 7). Congressional usage of agency performance reports was more positive than legislators' public rhetoric about the usefulness of the information in those reports. After beginning to receive

[31] The Senate Appropriations Committee (2000) noted, however, that the imperative of helping agencies learn to simply comply with the law detracted from substantive discussions on policy, and the strategic plans "…did not lay a good foundation for the performance plans" (3-4).

GPRA-mandated reports from federal agencies, the number of new laws and congressional reports citing GPRA increased steadily from 1997-2000, with 50 enacted measures in the 106th Congress (1999-2000) (U.S. Library of Congress, Congressional Research Service 2005, 3). Congressional Republicans continued to be GPRA advocates, but their concerns and disappointments with progress were revealing cracks in the legislative support edifice. Taking stock of agency compliance with GPRA with an obvious interest in its promise to eventually enable performance budgeting, congressman Stephen Horn, as committee chairperson, testified GPRA implementation was ". . .a work in progress" and indicated that, in 1999, GAO had determined ". . .only 14 of 35 Federal departments and agencies it examined were able to define some type of relationship between program activities on their proposed budgets and the performance goals cited in their plans" (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001, n.p.).

Senator Fred Thompson criticized the inadequate scope of performance goal setting for ". . .less than half of the critical management problems that the GAO and their IG's have identified," the reliability and timeliness of performance data, and the quality of agency performance reports" (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001). Senator Thompson also observed ". . .you can't really tell from the reports whether the agencies were making progress on many of their key outcomes. . ." and ". . .the performance reports don't tell decisionmakers or the public what's working and what's not within the mass of overlapping programs that exist in virtually every area of Federal activity." He conceded GPRA reporting had yet to be integrated into either the authorization or appropriations processes in Congress (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental

Relations 2001, n.p.). Gormley and Balla (2008) offer a likely explanation: ". . .congressional appropriators continue to pay little or no attention to GPRA reports when making resource allocations" (19).

In the Executive Branch, Deputy Director of OMB Sean O'Keefe admitted to a sense of discouragement, and assessed that GPRA was, by 2001, still far from achieving full acceptance and institutionalization within federal agencies:

> GPRA, with all due respect, has been treated by Federal agencies and departments, by and large as another reporting requirement; something else that needs to be complied with. And as a consequence of that, for it to be useful, for it to be really useful for any management purpose, it has to be introduced into the regular day-in-and-day-out management processes that are conducted throughout every Federal department and agency. It's nowhere near there. There are a very few interesting examples of how it's beginning to take hold, but those are noticeable by their distinguished nature of being so few, not because it's pervasive (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001, n.p.).

The GAO credited federal agencies with making "steady progress" in implementing GPRA through 2000, but found deficiencies in several areas: 1) the initial FY1999 and FY2000 performance plans contained few goals and measures that would allow assessment of program results, 2) agencies needed to engage in better coordination with other agencies on the development of performance goals and measures for co-managed programs in order to help address "mission fragmentation" and "program overlap," and 3) agencies needed to show clearer linkages between performance plans and budget requests in order to predict what impact budget changes would have on program performance (U.S. General Accounting Office 2000, 4-6). Additionally, A GAO analysis found limits on agency capabilities to perform program evaluations in the face of increasing demands: ". . . the resources allocated for conducting program evaluations were small and unevenly distributed across the 13 departments and 10 independent agencies we surveyed. . . ." (U.S. General Accounting Office 2000b, 11).

In 2001, John Mercer, as the former Senate Governmental Affairs Committee counsel, testified to Congress that his greatest disappointment with GPRA implementation had been that agency performance plans were being developed as the law required, but had not yet led to the development of performance plans at all agency levels, as he had envisioned (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001, n.p.). This helps explain why, as O'Keefe noted, GPRA was thus far irrelevant for routine agency management processes. GAO analysis of federal manager responses to its 2000 governmentwide survey offer empirical evidence to support the assessment that federal managers were generally not making use of performance information in routine management activities: over half of surveyed managers responded they used performance data for any of five categories of management activity to a "great" or "very great" extent in only 7 of 28 agencies (U.S. General Accounting Office 2001, 10).[32] Clearly, if some form of planning and performance management was occurring at lower levels of federal agencies, it was not yet systematically or comprehensively embedded in agency GPRA performance planning or evaluation routines.

If GPRA-style performance planning and measurement was to be institutionalized deeply into federal organizations, then planning within operating bureaus, divisions, and branches would first need to be integrated with GPRA-style planning that flowed downhill from departmental and agency strategic plans with their organizational goals and objectives. The "top down,"

[32] GAO noted three exceptions: over 50 percent of federal managers at HUD, VA, and GSA reported they used performance information for all five kinds of management activities GAO studied, which included setting program priorities, allocating resources, adopting new program approaches or changing working processes, coordinating program efforts with internal and external organizations, and setting individual job expectations (U.S. General Accounting Office 2001, 10-11).

deductive, and strategy-based planning perspective of GPRA needed to be merged with the inductive, "bottom up," and pragmatic planning framework used by program managers. In other words, budgeting and program evaluation needed to be integrated with strategic planning and organizational performance management. Beginning in January 2001, the George W. Bush Administration developed a President's Management Agenda (PMA) that, along with reforms in human capital management, e-government, competitive sourcing, and financial management, aimed to integrate performance and budgeting (Bouckaert and Halligan 2008, 390). According to OMB Director O'Keefe, the Bush Administration's top management priority was to require program budget requests beginning in FY 2003 to include performance measures and indicators (U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations 2001). To structure and evaluate agency inputs, it also mandated a new framework for program performance that would be tied to budgeting decisions it dubbed the Program Assessment Rating Tool, or PART.

Experimenting with Performance Budgeting with PART from 2001-2009

Unlike GPRA, which was a congressional initiative, PART originated in the Executive Branch under OMB championship and oversight (Gormley and Balla 2008). In the spring of 2002, OMB set up a task force with members from several OMB divisions to collaborate with stakeholders in GAO and the Congress to develop PART, and promulgated guidance to agencies that July to shape their inputs to a Performance and Management Assessments volume that would accompany the FY2004 President's Budget in February, 2003 and be posted to OMB's public website (U.S. General Accounting Office 2004, 62-63). To implement the Bush PMA and further drive performance management reform towards outcome-based performance budgeting, OMB replaced GPRA-mandated annual performance plans with annual performance budgets

starting in fiscal year (FY) 2005 through guidance to departments in OMB Circular A-11 (July 2003), part 6, section 200.1 (U.S. Office of Management and Budget 2003). Like performance plans, performance budgets described strategic goals and long-term and annual performance goals, each with associated targets expressed as outcomes[33] with their allocated resources (Bouckaert and Halligan 2008; U.S. Government Accountability Office 2004, 6). OMB took the position that ". . .performance budgets should satisfy all the statutory requirements of GPRA for annual performance plans" (U.S. Government Accountability Office 2004, 6).

PART was designed to be an evidence-based tool for OMB to rate program performance (U.S. General Accounting Office 2004, 5). (Kamensky 2021) argues PART shifted the unit of analysis from agency performance to major program-level performance because agency-level performance plans and measures were not useful to decision makers (n.p.). PART was a set of checklists with questions for agencies pertaining to selected programs' purpose and design, strategic planning, program management, and results and accountability that OMB examiners reviewed and scored (Bouckaert and Halligan 2008; U.S. Library of Congress, Congressional Research Service 2004). Program results were weighted at 50 percent in the OMB scoring model, with the combination of program purpose and design (20 percent), strategic planning (10 percent), and program management (20 percent) having an equal weight (U.S. General Accounting Office 2004, 9). OMB planned to evaluate as many as 1,000 major federal programs over 5 years (Gilmour 2007, 5).

PART ratings were intended as an input to OMB budgeting proposals for the President, and there is some evidence they mattered. While it did not yet include PART ratings, the

[33] Gilmour (2007) noted OMB made exceptions to the rule of measuring outcomes for research and development programs because outcomes for these programs, by definition, cannot be predicted *a priori* (12).

President's Budget for FY 2003, submitted in February 2002, reported the performance of more than 100 programs and was the first time a President's Budget had linked budget proposals with program performance evaluations (U.S. Library of Congress, Congressional Research Service 2005). The President's Budget for FY 2004 was the first to include PART ratings, but it only covered about 20 percent of federal programs that year, later increasing to about 80 percent by 2006 (Gilmour 2007). Programs in the FY 2004 budget request that had been rated "effective" or "moderately effective" received OMB recommendations for cumulative budget increases of 6 percent, and those rated as "ineffective" were recommended for increases of less than 1 percent (DeMaio 2003; U.S. General Accounting Office 2004a). However, statistically significant and positive correlations between PART scores and OMB changes to FY 2004 program budgets were only found in small discretionary programs (U.S. Library of Congress, Congressional Research Service 2004, 13). More convincing was that budgets of programs rated as "ineffective" were slashed by an average of 37.68 percent in the FY 2005 President's Budget (U.S. Library of Congress, Congressional Research Service 2004, 9).

Participant and observer experiences with PART, however, raised a host of concerns about the validity, reliability, and objectivity of its methodology. There was significant inconsistency in what OMB agreed to define as a "program" for PART reporting purposes due to levels of conceptual aggregation of government activities, how they were managed, and how they were resourced. In some cases, a program under PART could be a whole Bureau; in other cases, it could be a trifling scholarship program worth less than a million dollars (Gilmour 2007, 10). The Congressional Research Service (CRS) questioned the validity of program performance measures as a result of disagreements between OMB and agencies on appropriate performance goals, the reliability of performance measurement findings due to inconsistent

application of PART standards across programs, and the objectivity of data collection practices based on the subjective and ambiguous wording in some PART tool questions (U.S. Library of Congress, Congressional Research Service 2004, 16). GAO analysis of PART assessed that ". . . its effectiveness as a credible, objective assessment tool was challenged by inconsistency in OMB staff application of the guidance and limited availability of credible information on program results" (U.S. Government Accountability Office 2005, 5).

Inconsistency in the use of PART scoring standards may be explained, in part, by the fact that a single career official in OMB decided on the program score, and scores reflected political biases of their appointed superiors (Metzenbaum 2013, 858). Senior OMB officials at the time expressed dissatisfaction with the quality of the data and evidence of performance they received from agencies. For their part, many agency leaders struggled with evaluating the performance of their programs and feared their budgets would be unjustifiably reduced based on faulty or insufficient program performance evaluations. One example was the EPA, which struggled with measuring the effectiveness of Clean Air Act programs and was hindered by a 20-year decline in staff capacity for evaluating program performance from the 1970s to 1995, when its evaluation office was finally disbanded (Weinstock 2003, 40). The scope and purpose of evaluations, and the quality of evaluation designs, were topics of significant disagreement between agencies and OMB (U.S. Government Accountability Office 2005, 20-24). OMB's adverse assessment of agency program evaluation methods and goals was probably even more significant for program budgeting decisions than its assessments of agency performance reports on its programs, because the reliability and validity of the performance reports was subject to the perceived rigor of the agency's evaluation methods. In the case of agency evaluations of about 200 STEM-related programs, admittedly a difficult program category, former OMB Associate Director Robert Shea

indicated only 10 were considered minimally rigorous, and positive outcomes could be documented for only 3 (Heckman 2021).

Ironically, other PART observers found evidence that led them to question the reliability of PART performance measurements due to OMB's stated effort to *consistently* apply PART standards, which incorrectly assumed federal agencies were all Wilson Production-type organizations having programs with easily observable outcomes and outputs, without due regard for the reality that many federal organizations were of the Procedural, Craft, and Coping types, and so were not well situated to measure their program outcomes and/or outputs; this led to some agencies being penalized by OMB evaluators for having unmeasured program goals that were inherently unmeasurable and should not have been included in PART reviews (Gueorguieva et al. 2009, 229-230). Finally, PART and GPRA often used different indicators to measure the same programs, which led to contradictory performance ratings (Gueorguieva et al. 2009, 239).

Despite these criticisms of PART, the Director of OMB defended the evaluation program and believed it did more than just help OMB develop budgets: "Performance information gleaned from the PART…has also helped direct management, identified opportunities to improve program design, and promoted accountability. We believe that the PART will also greatly improve the goals and measures adopted through the GPRA strategic and performance planning processes" (U.S. General Accounting Office 2004, 66). GAO strongly rejected the basic premise of the latter claim:

> OMB's judgment about appropriate goals and measures is substituted for GPRA judgments based on a community of stakeholder interests. . . .GPRA is a broad legislative framework that was designed to be consultative with Congress and other stakeholders and allows for varying uses of performance information, while the PART applies evaluation information to support decisions and program reviews during the executive budget formulation process" (U.S. General Accounting Office 2004, 6-7).

The problem was that OMB's approach, designed to encourage use of performance information in the budgeting process, assumed pre-existing agency planning processes and capacity which did not yet exist, and to a great extent still do not. To truly complement GPRA, a PART-like system would have to depend on program-level performance measurement being conceptually and procedurally integrated with department or agency-level strategic planning goals and performance plans and measures. A bifurcated performance regime is a predictable outcome when agency strategic planning is not, as recommended by Mercer, institutionalized at all levels of the agency. Instead of replacing GPRA performance plans using PART criteria as OMB did, GAO recommended OMB foster a "complementary" relationship between GPRA and PART reporting routines (U.S. General Accounting Office 2004, 7). Of course, establishing complementary measurement systems would imply two different business processes that work with each other, so federal managers would have had to consider the resource implications for evaluation manpower and other capacities required for a two-level system, and the difficult cognitive work that would be demanded of managers to somehow integrate them. These costs and recognition of capacity shortfalls in federal agencies likely led OMB to simply replace GPRA performance plans with PART performance plans as a satisficing measure.

As they did during the early days of GPRA, legal and political institutional conditions mediated PART implementation. The politics and oversight conditions were complex. President Bush continued to serve as a champion for both management reform and performance throughout both terms (Breul and Kamensky 2008, 1015). In general, however, the Executive Branch was unable to engage congressmen or their staffs on management reforms (Breul and Kamensky 2008, 1020). In terms of implementation of PART for budgeting use, its primary purpose, Moynihan (2008) found that federal law and Congressional preferences required OMB

and federal agencies to report their budget data through a complex accounting framework that precluded clear association of performance goals and cost allocations with accounts reported to Congress (121). In addition, appropriations committee members and staffers were not receptive to the new changes in agency budget submissions that tried to incorporate program performance information, expressed a preference for pre-PART budget formats and, in some instances, even threatened to cut agency budgets if agencies continued to integrate performance data into budget documents (Moynihan 2008, 123-124).

A related critique of the use of budget lines for performance reporting under PART was that the unit of analysis, programs (or program activities – activities and projects funded by a budget line), was driven by budgeting categories that were often misaligned or not aligned to the organization and strategic plans of agencies (U.S. Library of Congress, Congressional Research Service 2004, 5). In addition, other forms of legal constraints abounded. In some cases, the program outcomes or outputs PART tried to measure were disconnected from Congress' intent for the program. Specifically, efforts to implement PART that were focused on improving the efficiency of programs conflicted with legal mandates on agencies designed to achieve legislated and policy goals unrelated to efficiency. For instance, the legal mandate for the Federal Electoral Commission's Compliance and Enforcement Program to make decisions in a way that restricted partisanship met with criticism from OMB PART evaluators because, in a very few cases, the legislated business processes preventing partisanship led to an outcome OMB deemed as inefficient (i.e. an impasse) (Gueorguieva et al. 2009, 239).

As with many performance evaluation regimes, there was a high risk of politicization in PART evaluations and associated budget decisions. In 2004, the CRS warned "subjectivity in completing the PART questionnaire and determining PART scores could potentially. . .be

introduced by White House, OMB, and other political appointees" (U.S. Library of Congress, Congressional Research Service 2004, 8). CRS was prescient. PART evaluations were found to be more important for OMB budget decisions concerning programs in departments traditionally supporting Democratic agendas than they were for programs in departments implementing Republican priorities, resulting in lower PART scores and smaller budget increases on average for programs favored by Democrats, when controlling for other factors (Gilmour and Lewis 2006, 750-751). It seems plausible that if the Obama Administration had continued using the PART process, PART would have mattered more for Republican-oriented programs. The PMA, of which PART was but one element, and many other political agendas and policy interests of the Administration[34] exerted influence on budgetary priorities, thereby limiting the role played by performance information in budgetary deliberations and OMB budgetary recommendations.

More so than GPRA, PART serves as a useful case study on the impact of resource dependency on agency implementation of reform, and how it combines with real or perceived political influence and ideology to motivate action by federal managers. As OMB implemented PART, it informed departments and agencies that they risked administrative budget cuts if too many of their programs were designated as "results not demonstrated" due to inadequate PART performance measures; this veiled threat led to a noteworthy drop in programs without acceptable measures from 98 in FY 2004 to 36 in FY 2007, and created incentives for federal

[34] Congress and the Administration were not the only actors with policy interests. According to Gilmour (2007), the Department of Education intentionally refrained from seeking to improve its very low program effectiveness PART ratings in order to bring attention to what it believed were many poorly-designed programs it was tasked to administer (17). In this case, the absence of leadership commitment to performance management reform was due to a higher commitment to policy reform for educational programs.

program managers to take PART more seriously based on their perception that OMB meant business and would continue to use PART to inform budgetary decisions (Gilmour 2007, 13).

Notwithstanding this research, there is persuasive evidence that the incentive effect varied between federal agencies depending on the degree of divergence of their organizational ideology with that of the Administration and the self-identification of political party by federal managers. For instance, federal managers, regardless of party affiliation, who worked in agencies that managed more "liberal" programs perceived a greater administrative burden from PART than did managers in agencies with more "conservative" programs, even when controlling for the types of programs their agencies managed (e.g. block grants, research and development, or direct delivery) and their evaluative capacities (Lavertu, Lewis, and Moynihan 2013, 851, 854). Conversely, managers who self-identified as Democrats, regardless of agency, also perceived higher levels of administrative burden in their agencies on average than Republican managers did (Lavertu, Lewis, and Moynihan 2013, 854). The effect of these perceptions of heavy burden suggests ideological differences may have negatively mediated the effects of management efforts to institutionalize performance information uses in "liberal" agencies (Stephane Lavertu and Moynihan 2013). The history of political influences during the PART era therefore continued the pattern of political influence being both a contributor and a detractor to reform institutionalization.

During this period, the effects of senior leadership commitment continued to measurably affect the implementation of PART. Even when agencies had programs whose outcomes could be difficult to quantify and measure, strong and credible leadership by political appointees communicated the degree of importance and seriousness with which those leaders intended to implement PART. For example, managers at the State Department uniformly credited Secretary

of State Colin Powell and his Deputy Richard Armitage for investing unaccustomed levels of attention on PART and other internal management issues, helping State to lead all other agencies in the percentage of programs rated as "effective" and the second lowest number of programs receiving a "results not demonstrated" rating by FY2007 (Gilmour 2007, 15). Dull (2009) argues that, in the public sector, credible leadership commitments depend on the "reputational credibility" of senior leaders (273). In this case, Powell's impressive reputation as a national leader and a highly-effective manager preceded him into office and motivated managers throughout State to implement PART after he communicated it was important to him.

Also, during this period, Congress and the Office of Personnel Management (OPM) revised human capital rules to strengthen incentives for credible commitment by senior executives to actively manage agency performance. Agency executive leadership commitment to performance management started to be fiscally incentivized after reforms to senior executive compensation practices in the FY 2004 National Defense Authorization Act (NDAA) (National Defense Authorization Act 2004, Pub. L. No. 108-136, 117 Stat, 1638 (Nov, 24, 2003)) that directly tied executive compensation to how well individual performance supported agency strategic goals. GAO assessed this as "…a substantive and positive step for agencies in holding senior executives accountable for their performance and contributions to organizational success" (U.S. Government Accountability Office 2006, 4). Findings from a 2008 GAO survey of federal managers suggests executive perspectives on their accountability for results improved between the FY2004 NDAA and 2007, whereas the views of non-SES managers not covered by the NDAA changes remained the same:

> . . .there was a 14 percentage point increase in the number of SES who responded that managers/supervisors at their level are held accountable for accomplishment of agency strategic goals. In 2007, there was a 12 percentage point increase in the number of SES who reported that they are held accountable for the results of the programs, operations, or

> projects for which they are responsible as compared to 2003. . . .There was no significant change in responses from 2003 to 2007 in non-SES level responses to either of these questions (U.S. Government Accountability Office 2008, 13).

Based on the GAO 2007 survey of federal managers, statistical analysis suggests that by the end of the Bush era, federal managers broadly perceived that senior leadership commitment to results positively influenced the development of performance measures and goals and the purposeful use of performance information for a wide variety of management purposes, including program management, employee management, and problem solving (Moynihan and Lavertu 2012, 599).

Department and agency cultures continued to evolve towards a results orientation, but there was much greater progress by agencies in adopting performance measures than in using performance information for routine management decisions. Using 2007 GAO survey data from over 2,000 federal manages, Moynihan and Lavertu (2012) found involvement in PART tasks only improved agency passive use of performance information for performance measurement tasks (goal setting and performance measurement) and, except possibly for informing process changes, not for routine use for management decisions (598).

Practice-driven institutionalism may help explain this divergence. It postulates institutional agents in organizations "reproduce and modify" their collective "understandings" from practice and then use this knowledge to decide where, when, and how to enact changes to best cope with the complexities of multiple and often competing institutional logics (Smets et al. 2015 as cited by Smets, Aristidou, and Whittington 2017, 380). Such enactments by federal managers should contribute to the institutionalization of a results orientation in managerial attitudes and the development of business practices that use performance information whenever managers accord sufficient priority to the logic of performance management versus potentially competing logics, such as professionalism, innovation, or citizen engagement. The evidence leads to the conclusion that, by the end of the Bush Administration, most federal managers were

not yet giving the logic of performance management the same priority as prevailing institutional logics due to concerns about losing legitimacy with what Smets et al. (2015) refer to as a "referent audience" (934) backing the prevailing logics. In other words, cultural norms and values in most federal agencies still led managers to view GPRA and PART as their own set of rules and requirements, not a tool to add to other existing ones for creating value for management decisions or agency performance. The start of institutionalization of performance management in daily operations would have to wait for cultural changes in the next phase of reform in which managers would begin to perceive performance information as a useful complement to professional judgment and other management strategies they used to drive high performance.

Notwithstanding the immature state of federal manager use of performance information during the PART regime and OMB testimony, empirical research on federal performance management reform revealed evidence some federal program managers were using performance information to make important program decisions. A study by the IBM Center for the Business of Government in 2003 acknowledged federal program managers had "…limited experience to date in outcome measurement and in the use of outcome data," but argued progress was being made (Hatry et al. 2003, 12, 18). The IBM study cited 16 case studies from federal departments in which program managers, using more routinely-available performance data on program outcomes, were able to put it to use to take corrective action, identify best practices, motivate staff, and plan and budget. A straightforward example was the Department of Education's use of state-provided outcome information on adult education and literacy outcomes from federal grant programs to inform management decisions on locations for staff assistance visits and for targeting of technical assistance and training (Hatry et al. 2003, 21-22). Another example in the

IBM report was the Environmental Protection Agency's use of outcome data on different kinds of air pollution to propose legislation; create or change standards, rules or regulations; take corrective actions, and track improvement (31).

A content analysis of the IBM report shows two of the cases pre-date GPRA, 11 began measurement practices during the GPRA era, and three did not report a start date for measurement but likely pre-dated GPRA and simply exemplified longstanding "watchful eye" compliance-related collection practices (such as EPA air quality measurements, Occupational Safety and Health Administration inspections, and Social Security disability fraud monitoring). Four cases indicate some association of outcome measurement with GPRA, and only one, the Coast Guard's Maritime Safety, Security and Environmental Protection program, explicitly responded to GPRA outcome measurement philosophy. In terms of manager use of outcome data for decision making, the examples in the report show a trend towards management use of outcome information a few years after the start of program outcome measurements, with six programs starting use for decisions during the GPRA era and seven during the PART era (and three unknown). This suggests outcome-based performance measurement practices generally began to take root during the GPRA period, but most agencies required several years to institutionalize new practices for outcome performance data to be used in management decisions during the later GPRA and PART eras. Therefore, shortly after the start of the PART period, it is fair to conclude institutionalization of practices for measuring and using outcome data in federal agencies, a core goal of liberation management reform, was still at the pre-habitualization stage and not specifically tied to GPRA or PART.

Near the end of the Bush Administration, lessons learned from the GPRA and PART experiences led to changes in how performance management would be governed during the

Obama Administration. Executive Order (EO) 13450 (November 13, 2007) strengthened the institutional governance of GPRA and PART by establishing a Performance Improvement Council (PIC) and Performance Improvement Officer (PIOs) of senior executive rank at all CFO Act agencies (U.S. President 2007). Senior leader commitment to goals was hindered by the amount of performance data generated by both GPRA and PART, so the next period would reduce the number of goals substantially to allow leaders to focus on a manageable number (Moynihan and Lavertu 2012, 600). Congressional staff reported similar challenges finding performance information they thought was relevant; also, federal staff members indicated GPRA and PART had shifted the performance focus of agencies towards the measurement of outcomes vice outputs, but reported agency workloads had significantly increased, there were problems due to lack of linkage between GPRA and PART, subjectivity in PART measures was continuing, and there was an absence of processes for resolving agency disputes with OMB over program measures, targets, and evaluations, among other unresolved implementation issues (Metzenbaum 2009, 25). A more positive legacy of PART was, as noted by GAO, an increase in the attention of agencies to the importance of building capacity for evaluation and making program performance information available (U.S. Government Accountability Office 2005, 28). Perhaps of ultimate concern for governance was the perception that PART had undermined the legislative purposes for which GPRA had been designed by causing goal displacement in federal agencies: "…OMB's emphasis on PART caused agencies to devote their time to earning a good PART score rather than to finding ways to improve outcomes, return on spending, and the quality of people's interaction with government" (Metzenbaum 2015, 46). Such candid feedback would frame leader and stakeholder dialogue about reforming the reform once again at the beginning of the Obama Administration.

Strengthening Institutional Governance with GPRAMA from 2010 to Present

After the start of the new Obama Administration in January 2009, Congress quickly reasserted its leadership of performance management reform, and movement to improve its institutionalization within Congress was evident. A bipartisan Task Force on Government Performance was established under the Senate Budget Committee. Democratic Senator Mark Warner of Virginia was appointed as Chairman based on his leadership of performance reforms in Virginia State government while he was governor and his admitted "obsession" with performance measurement (U.S. Congress. Senate. Committee on the Budget. 2009, 3). This effectively situated him as the performance movement's next heir apparent thought leader in Congress, bringing innovative practices with him from Virginia, much as John Mercer had been early in the GPRA era coming from California. Reflecting the bipartisan nature of the support, Republican Senator Jim Bunning of Kentucky in the Senate Budget Committee expressed support for the new task force based on the traditional Republican focus on saving money to help manage skyrocketing federal debt loads (U.S. Congress. Senate. Committee on the Budget. 2009). Thus, the long legacy of bipartisan congressional support for performance management survived yet another Presidential transition despite evidence of its politicized use against programs favored by Democrats during the previous administration. An early finding of the task force was that Congress, OMB, and agencies were not yet making full use of GPRA data to inform decisions, so structuring GPRAMA, or the "Modernization Act," to encourage greater performance information use became a high priority (Amy Edwards, as cited by Moynihan (2013, 11)).

The new Administration's rejection of the PART approach likely facilitated Congress' reassertion of its authority. In testimony before the Senate Budget Committee, the new OMB

Deputy Director of Management and first Chief Performance Officer (CPO), Jeffery Zients, criticized the PART system as a failure. He believed it was unused by Congress and agencies and produced vast amounts of information that was not meaningful to the public, then declared his intention to replace it with a new performance evaluation system. Zients promised the new system would 1) focus agency efforts on a much smaller number of significant goals and would demand senior leader ownership, 2) cascade planning goals from agency down to program and individual levels, 3) develop outcome-oriented cross-agency goals with accountable "goal owners," 4) demand more frequent "relentless" review of progress at program, unit, and agency levels towards goals to embed performance measurement routines deeper into agency cultures, and 5) require greater transparency through engagement with Congress, the public, and the federal workforce (U.S. Congress. Senate. Committee on the Budget. 2009, 10-11, 15). Zients appointed Shelley Metzenbaum, on the basis of a report she had written for the IBM Center for the Business of Government, to lead the new Administration's performance reform efforts in OMB. Metzenbaum (2015) described the Obama Administration's new philosophy:

> The Obama Administration aimed to reset the performance mindset of US federal agencies. We wanted people across the US federal government focused on improving performance, not on getting higher ratings from budget examiners, attaining a higher percentage of targets met, or complying with reporting requirements for the sake of compliance (45).

Indeed, this approach acknowledged that, despite its sharp focus on measuring program-level outcomes, PART had not improved performance management reform institutionalization in the Executive Branch beyond refinements in program measurements and use of performance information for OMB-level program budget reviews, which the previous section showed were often viewed as subjective, invalid, unreliable, and in some cases politicized. GAO found that ". . .while managers have a lot more information today than they did 10 years ago. . .there really

has not been much progress over that period in the extent to which they use information to make decisions that could lead to better results" (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 7).[35] GAO investigations of several federal agencies reporting relatively low and high usage of performance information revealed that several conditions determined the likelihood of agency use of performance information: credible leadership commitment to achieving results, the quality of performance information for managerial decisions (a function of performance measurement system maturity), and managerial attitudes towards performance information generally (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 7-8). The availability of training on the use of performance information was, nevertheless, identified by GAO as the single most important predictor of managerial information use historically (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 13).

Congress wasted no time introducing new legislation to reform GPRA and retain the few portions of PART that had withstood critical review by OMB, the Performance Improvement Council (PIC), and the good government community. Perhaps the most accepted PART practice

[35] Steinhardt noted there were exceptions in which performance information had been used by managers and resulted in improved outcomes, such as reduced highway fatalities as a result of the National Highway Traffic Safety Administration's "Click it or Ticket" program which encouraged drivers to wear seat belts (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 13). These were clearly exceptions that proved the rule.

that carried over to GPRAMA was frequent but selective reexamination of federal programs, according to Paul Posner (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 26). Representative Henry Cuellar from Texas, praising effective GAO research support, introduced H.R. 2142, initially titled the "Government Efficiency, Effectiveness, and Performance Improvement Act," on June 16, 2009 and testified to the Subcommittee on Government Management, Organization and Procurement of the House Committee on Oversight and Government Reform that the bill would require all federal programs be reviewed every five years, require agencies to submit program assessment reports and management improvement plans to address issues spotlighted in their assessment reports, and make the PIC and agency Performance Improvement Officers (PIOs) statutory roles (U.S. Congress. House. Committee on Oversight and Government Reform. Subcommittee on Government Management, Organization, and Procurement 2009, 61-65). The Democratic-led Committee on Oversight and Government Reform approved H.R. 2142 by voice vote on May 20, 2010 and the full House passed it by voice vote on June 6, 2010. In the Senate, according to Kamensky (2021), the bill was amended to the point of a near total re-write using language developed in OMB that broke decisively away from PART philosophy in favor of a significantly refined GPRA approach (interview with author). The Senate passed it by unanimous consent on December 16, 2010.[36] After heated debate between representatives Cuellar and Issa, who was

[36] Representative Cuellar, speaking on HR 2142, 111th Cong., 2d sess., *Congressional Record* 172 (December 21, 2010): H 8857-8858. https://www.congress.gov/111/crec/2010/12/21/CREC-2010-12-21-pt1-PgH8852-5.pdf

opposed to the Senate-amended version of the House bill,[37] the House, doubtlessly considering the change of party leadership from Democrats to Republicans looming in the House the following month which would likely derail the legislation, voted in favor during the final hours of the 111th Congress. President Obama signed GPRAMA into law on January 4, 2011 (figure 3.1).[38]

Figure 3.1. President Obama signs GPRAMA into Law (January 4th, 2011)

Source: https://obamawhitehouse.archives.gov/omb/performance/president-signs-gprma. Standing behind President Obama are then-former Acting OMB Director Jeffery Zients and Associate OMB Director for Performance and Personnel Management Shelley Metzenbaum.

GPRAMA established significant new planning, reporting, and oversight requirements for agencies and the federal bureaucracy at large, established new leadership roles, gave OMB additional authorities and responsibilities, and strengthened rules for system-wide accountability

[37] *GPRA Modernization Act of 2010,* HR 2142, 111th Cong., 2d sess., *Congressional Record* 172 (December 21, 2010): H 8860-8861. https://www.congress.gov/111/crec/2010/12/21/CREC-2010-12-21-pt1-PgH8852-5.pdf

[38] *GPRA Modernization Act of 2010*, Pub. L. No. 111-352, §1, 124 Stat. 3866 (2011). https://www.govinfo.gov/content/pkg/STATUTE-124/pdf/STATUTE-124-Pg3866.pdf

through multiple improvements to governance and executive leadership. Comparing the sheer length of the GPRAMA and the GPRA laws provides a first clue as the magnitude of the changes which incorporated multiple new provisions and enhanced existing ones: GPRA contained 4,742 words, while GPRAMA had 7,410. Appendix A compares the GPRA, PART, and GPRAMA performance regimes in terms of their products and processes categorized by their associated federal and agency levels and by their respective mechanisms for accountability, budgeting integration, Congressional oversight, and transparency to citizens. Appendix B presents how the various products and processes work together and are sequenced today at the federal and agency levels to support planning, the evaluation and analysis of evidence, and reporting. Note that the performance management framework at appendix B includes additional elements such as agency strategic objectives, learning agendas, evaluation plans and enterprise risk management assessments that have been added by OMB in response to related legislation after GPRAMA was signed that were designed to ensure more comprehensive use of data, evidence, and risk management methods in the performance management process.

Lessons learned from GPRA and PART implementation informed new GPRAMA requirements. Moynihan (2013) argued the GPRAMA enhancements were designed to:

> . . .generate more timely data and push leaders to articulate their priority goals more clearly. The new roles create a team of senior managers with the job of leading the performance system, setting goals, ensuring information is available, and conducting frequent data-driven reviews. The new processes seek to structure ongoing attention to the analysis of performance data for improved execution and strategic decisions and for better communication to delivery partners and other stakeholders (9).

From GPRA and PART, the gained insight that the performance of many essential government programs was impossible to validly measure at agency level because they were orchestrated across multiple agencies using diverse strategies, policies and agency program activities led to the requirement for a new government-wide tier of performance management actively managed

by OMB. Arguably, the most substantial expansion in the scope of the performance management reform instituted by GPRAMA was the introduction of cross-agency priority (CAP) goals and federal government performance plans, which included goals for federal policy outcomes and for improving internal management processes in agencies. This expansion created a tool to enable Congress and the President to evaluate the performance of cross-agency programs and identify opportunities to reduce ". . .fragmentation, overlap, and duplication."[39] Along with an enlargement of OMB's role for coordinating and evaluating cross-agency progress towards CAP goals came a new OMB leadership philosophy. Even before GPRAMA was signed, the new CPO sought to transform the role of OMB from a compliance enforcer to serving as a "teacher," a "facilitator," and "a hub of best practice sharing" through the Performance Improvement Council (PIC) (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, n.p.)

Limited capacity for planning and performance evaluation, especially managerial time and attention, was similarly acknowledged as a key predictor of agency responsiveness to reform, and its effects inspired proposals to limit the number of goals managers would have to focus on concurrently. Whereas no such distinctions were made in GPRA or PART, mandating

[39] *Government Efficiency and Effectiveness: Strategy for Reducing Fragmentation, Overlap and Duplication and Achieving Cost Savings: Testimony before the Committee of the Budget, U.S. Senate*, 113th Cong. 1 (2013) (statement of Gene L. Dodaro, Comptroller General of the United States)

the designation of selected goals at federal[40] and agency[41] levels as "priority" permitted sufficient allocation of scarce federal manager attention to a limited number of goals in order to increase performance information use and the allocation of resources related to those goals (Moynihan 2013, 12). Combined with the quarterly frequency of progress reviews of CAP and agency priority goals, the law anticipated the new approach to selective focus on priority goals would help push agency leaders to embrace the reform and move beyond a simple attitude of compliance in their responses to the reform (U.S. Government Accountability Office 2011, 9).

Prioritizing goals was not the only strategy GPRAMA used to try to institutionalize performance management beyond simple agency compliance. Commanding greater attention from political appointees to focus them on performance management was essential as well. According to Jeffery Zients, OMB Deputy Director for Management and the first Chief Performance Officer, testifying before a Senate subcommittee in 2011:

> I think, historically, one of the problems had been the turnover and lack of focus of political folks, and we have been working very hard to make sure that deputy secretaries and senior teams are focused on a handful of priorities…we tend to measure everything, and then nothing really matters. So I think central to the Modernization Act. . .is to pick

[40] OMB Circular A-11, part 6, defines a CAP goal as "a statement of the long-term level of desired performance improvement for Government-wide goals set or revised at least every four years. These include outcome-oriented goals that cover a limited number of crosscutting policy areas and management goals addressing financial management, strategic human capital management, information technology management, procurement and acquisition management, and real property management" (page 19 of section 200). https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf

[41] OMB Circular A-11, part 6, indicates agencies should carry "…a limited number of goals, usually 2–8, identified by CFO Act agencies or as directed by OMB. An APG [Agency Priority Goal] advances progress toward longer-term, outcome-focused goals in the agency's Strategic Plan, near-term outcomes, improvements in customer responsiveness, or efficiencies. An APG is a near-term result or achievement that leadership wants to accomplish within approximately 24 months that relies predominantly on agency implementation (as opposed to budget or legislative accomplishments). APGs reflect the top near-term performance improvement priorities of agency leadership, not the full scope of the agency mission" (page 20 of section 200). https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf

> a number of priorities and track them relentlessly" (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 15).

The tacit implication of this insight, of course, is that the system of political appointments to federal agencies represents a systemic limitation on agency capacity to analyze performance information and use it for management decisions because it reduces the number of goals agency leaders can functionally attend to at any given time. To mitigate this, GPRAMA institutionalized official performance management titles and augmented role descriptions for key executives in CFO Act agencies in order to improve their accountability for their agency's performance and results (U.S. Government Accountability Office 2011, 9-10; 2013, 3). Whether intentional or not, this governance change built on the earlier NDAA senior executive performance-based financial incentives tied to their contributions to agency strategic goals discussed previously.

The new roles included the appointment of the OMB Deputy Director for Management as the CPO,[42] the legal designation of PIOs, and the establishment of the PIC chaired by the CPO. In agencies, Deputy Agency Heads were designated as the Chief Operating Officer (COO) and authorized to oversee the efforts of the PIO, Goal Leaders, and internal agency performance measurement activities. Jonathan Breul, then Executive Director for the IBM Center for the Business of Government, testified in 2011 that "the COO position…elevates management to a level where both policy and management report. It brings them together where they can really get traction in the department. And this new provision in law provides both an incentive and a discipline to seek people with the right qualifications for the job" (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 31). OMB later reinforced

[42] The CPO role was not a part of GPRAMA. Kamensky (2021) stated it had been an Obama campaign pledge that received little support from the interagency community, so the White House opted to dual-hat the OMB Deputy Director for Management as the CPO in order to create an appearance of fulfilling his campaign promise (interview with author).

the need for sustained effort and consistency in federal agencies by directing agencies with a politically-appointed PIO or a PIO with limited appointment period to select a Deputy PIO, and to appoint Deputy Goal Leaders to support Goal Leaders (where the Goal Leader is a political appointee, OMB encouraged agencies to select a career senior executive to fill the role) (U.S. Government Accountability Office 2013, 4-5).

Technical refinements in the statute were also needed to strengthen the timing and integration of the new performance management framework with federal and agency agenda setting and strategic planning processes. To improve the timing alignment of agency strategic goal development with the PMA, 5-year agency strategic plan updates were changed to coincide with 4-year presidential terms. Additionally, a clear description of how agency performance goals contribute to CAP goals was mandated, and the requirement for a close linkage of agency performance goals with agency strategic plans was made more explicit. The tempo of agency priority goal progress reviews under GPRAMA extended the applauded PART routine of conducting program reviews quarterly, a standard to which OMB was also held for CAP goals. Should an agency fail to meet one of its strategic goals, GPRAMA provided that OMB would review the circumstances and determine appropriate actions to remediate the deficiencies, increase budget obligations, or even, after several years of unmet goals, propose changes in statutes or program termination. The potential of the Internet to improve transparency was leveraged as well by converting an existing OMB-managed government intranet site (performance.gov) for use in posting government-wide and agency plans and performance reports to the public Internet (Moynihan 2013, 15).

With this portrayal of the establishment of and actions to implement GPRAMA, we can now turn to sketch, in broad form, the institutionalization outcomes of the reform, and then,

using the best available evidence, how the government-wide institutional and organizational conditions in the inferential model influenced them. These descriptions will serve as the historical context for the subsequent mixed methods analysis of data from 2017 and 2021 that seek to identify which configuration of conditions, by their presence or absence in federal agencies, best account for agency institutionalization of GPRAMA.

As described in chapter 2, this study conceptualizes the enactment phase of reform policy implementation occurring under the control of managers in the action situation of the inferential model. In this model, based on the public management literature, GPRAMA institutionalization is the combination of two measurable indicators: the adoption of performance measures and the use of performance information for management decisions. In the context of the preceding history, it should be quite clear that Congress, most stakeholders, and reformers were strongly interested in agencies adopting outcome-oriented performance measures to instill a result-oriented culture in agencies, and then using the useful and high-quality performance data routinely to make important management decisions about program plans and operations to improve agency performance. These, then, are the historical indicators to evaluate the degree of GPRAMA institutionalization government-wide and inside federal agencies.

Adopting meaningful performance measures continued to challenge federal agencies during the GPRAMA period, but some progress was made as agency strategic planners became more adept at articulating outcome-oriented strategic objectives as part of agency strategic plans, and performance management routines became more embedded in agencies and familiar to managers. Stanley and Lutz (2021) compared federal agency strategic plans in 2013 and 2017, and found that the percentage of strategic objectives with mission-related external outcomes increased from 16 to 26 percent over the period. Across agencies, they reported that “in 2013,

more than 50% of the strategic objectives were externally-oriented in all but 7 agencies…in 2017, at least 50 % of the strategic objectives were externally-oriented in 18 of 19 agencies for which comparisons could be made" (15). Comparing federal manager responses of agreement to a "great" or "very great" extent that they had difficulty determining meaningful measures for their programs, operations or projects they managed, the governmentwide percentage fell from 30.0 percent in 2013 to 28.5 percent in 2017 (U.S. Government Accountability Office 2013a; 2017c).

Developing and using measurable outcome-focused performance goals remained the most elusive requirement for many agencies, but some agencies were able to advance beyond simply measuring their processes and outputs. In 2015, GAO found that "…agencies continue to make insufficient progress in establishing and using outcome-oriented performance measures," singled out as challenged programs the Department of Defense (DoD) and Department of Veterans Affairs (VA) joint project to make veteran health records interoperable, the Department of Justice (DOJ) program to address incarceration challenges, and DoD efforts to reduce incidents of sexual assault in the military, and recommended OMB facilitate sharing of measurement best practices between Goal Leaders (U.S. Government Accountability Office 2015, 58). In some cases, nevertheless, early agency challenges in defining outcome measures led to constructive reflection, learning, and then the development of new priority goals, as with the Department of Agriculture when senior leaders discovered the department lacked a method to evaluate the effects of program measures on water quality improvement (an outcome), then set a priority goal to develop these methods (Mills 2013 as cited by Metzenbaum (2015), 54). Other agencies were initially prevented from adopting outcome measures due to the lack of supporting software systems or business processes. One example was OPM's lack of an inventory

management system to track the age of pending retirement claims, but it eventually innovated a strategy to measure customer wait times (Metzenbaum 2015, 54). Important dimensions of government-wide federal performance measurement analysis did not make a substantial improvement, if federal manager viewpoints serve as evidence. GAO 2013 and 2017 federal manager survey data (table 3.2) showed a slight improvement in the availability of performance data for managers and employees, but little in other studied indicators.

Table 3.2. Indicators of Performance Measure Analysis in CFO Act Agencies (2013 and 2017).

Question from GAO Survey of Federal Managers	2013	2017	Net Change
Data Accessibility Indicators			
My agency's performance information is easily accessible to managers at my level.	48.7%	50.6%	1.9%
My agency's performance information is easily accessible to employees.	33.6%	35.7%	2.1%
I have access to the performance information I need to manage the program(s)/operation(s)/project(s) that I am involved with	51.7%	51.7%	0.0%
Data Transparency Indicators			
My agency's performance information is easily accessible to the public, as appropriate.	17.1%	17.0%	-0.1%
Useful Formats for Managerial Use			
My agency's performance information is available in a format that is easy to use.	32.1%	32.1%	0.0%
Timeliness of Information			
Performance information is available in time to manage the program(s)/operation(s)/project(s) that I am involved with	44.2%	42.8%	-1.4%

Notes

1. Data shows percentage of managers responding "to a great extent" or "to a very great extent"

Sources of data: U.S. Government Accountability Office. 2013. “Managing for Results: 2013 Federal Managers Survey on Organizational Performance and Management Issues,” GAO-13-519SP (June 26, 2013). https://files.gao.gov/special.pubs/gao-13-519sp/results.htm#question_72; U.S. Government Accountability Office. 2017. "Supplemental Material for GAO-17-775: 2017 Survey of Federal Managers on Organizational Performance

and Management Issues," GAO-18-776SP (Sep 29, 2017). https://files.gao.gov/special.pubs/gao-17-776sp/resultsall.htm

The question of whether federal managers made greater use of performance information, despite the greater volume and quality of performance data, is subject to considerable dispute. Moynihan and Kroll (2016) found evidence that federal agencies responding to GPRAMA requirements had indeed established new routines that contributed to greater managerial use of performance information (320). A leading good government observer, author, and advocate, John Kamensky of the IBM Center for the Business of Government, supports Moynihan and Kroll's assessment (J. Kamensky 2021). When examining the institutionalization of specific kinds of routines, GAO found contrary evidence concerning data-driven routines in agencies for reviewing performance which sheds light on the shallowness of organizational institutionalization of performance goal setting: "GPRAMA requires agencies to conduct . . .reviews for agency priority goals, which represent a small subset of goals, and they are required at the departmental level. These reasons may explain why most managers reported they were not familiar with the reviews" (U.S. Government Accountability Office 2018, 36). In this example, greater reliance on priority goals focused management attention on GPRAMA requirements at the top levels of departments and agencies, perhaps inhibiting institutionalization of GPRAMA goal-setting routines into program and street levels. Perhaps more tellingly, GAO found that government-wide, there was no statistically significant difference in average manager use of performance information across federal agencies from 2013-2017 (U.S. Government Accountability Office 2018, 14-15). Since these managers were only lightly involved in goal-setting upfront, it makes sense they would not make use of performance information generated on agency goals with which they were not generally familiar.

Political conditions during the implementation of GPRAMA can perhaps best be described as a case of benign neglect. At the dawn of the GPRAMA era in 2009, Professor Paul Posner at George Mason University testified in a Senate hearing that he noted continuity in the reform with the new Administration choosing to sustain it, and expressed his opinion that " . . .the use of performance in authorizing and appropriations committees is actually fairly impressive" (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 25-26). It can be argued that the political conditions were benign only because there was little overt opposition to its substantive requirements for OMB and federal agencies, but offered potential payoffs for legislators who could use their engagement in the reform to present themselves as good fiscal stewards during a time when mounting federal deficits and debt were leading many to call for greater transparency and accountability for programming and budgeting decisions. It was therefore framed by congressional leaders as an issue of good government seeking to assure the public that their government was making efficient use of tax dollars, and not grounds for partisanship (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 11). Congressional interest in GPRAMA, to the extent it existed, was motivated more by a pursuit of narrow political agendas than of higher levels of government performance. For example, during a committee hearing on the new law shortly after it was signed, committee chairperson Senator Thomas Carper (D-Delaware) expressed his hope that OMB would use GPRAMA as a tool to help inform complex decisions on merging or reorganizing agencies (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 3).

By the fall of 2012, a conference sponsored by the National Academy for Public Administration and the IBM Center for the Business of Government with broad participation by members of Congressional staffs, OMB, and federal agencies found that, while GPRAMA mandated improved agency coordination with Congress, most members of Congress were uninformed about it or had little incentive to be interested in performance information and were "ambivalent" about engaging with it (Moynihan 2013, 30). By 2015, the Republican-led House Committee on Oversight and Government Reform intended to more closely examine OMB and agency progress in implementing the GPRAMA requirement to publish a complete inventory of federal programs, noting that ". . .many federal programs are not identifiable and the data is not comparable, making the program inventory nearly unusable" (U.S. Congress. House. Committee on Oversight and Government Reform 2015, 168).[43] Therefore, while Congress continued to monitor GPRAMA implementation to some degree, interest in results-based reform was limited to a handful of congressmen and senators and their interest was driven more by perceived opportunities to reduce duplication and overlap between programs in order to save money than it was on improving program performance or agency use of performance information to improve management effectiveness. Evidence for this claim can be seen in federal manager perspectives on the low level of congressional interest in their use of

[43] In Apr 2013, Rep. James Lankford (R-Oklahoma) introduced H.R. 1423, the *Taxpayers's Right-to-Know Act* (https://www.govtrack.us/congress/bills/113/hr1423). The bill was designed to amend (expand) GPRAMA reporting requirements regarding agency programs, including reporting administrative costs, expenditures for services, number of clients served and beneficiaries who received assistance, the number of full-time federal employees who administer the program, the number of employees paid for by the federal government through a grant, contract or agreement, identification of the statute authorizing the program, and any finding of programmatic duplication or overlap as reported by internal review, an inspector general, GAO, or other source, and program performance reviews including as required by GPRAMA (H.R. 1423, 113th Cong., 2d sess., 2014).

performance information. In 2013, only 17.2 percent of federal managers agreed to a "great" or "very great" extent that members of Congress or their staffs paid attention to agency use of performance information for management decisions; in 2017, it was 17.1 percent, and in both cases over 40 percent of managers responded they had no basis to judge or the question was not applicable – indicating to them that congressional oversight was not salient (U.S. Government Accountability Office 2013a; 2017b).

Federal agencies adapted to new GPRAMA requirements for performance goal setting. By requiring the development of new CAP goals and agency priority goals, GPRAMA forced federal leaders to decide which crosscutting and agency goals were truly the highest priority for near-term action. Isolating priority goals should therefore help reduce overall goal ambiguity and provide a sharper focus on priority goals and an opportunity to more cleanly differentiate lower-priority agency goals from them. In 2013, during a hearing of the Task Force on Government Performance, Senator Warner observed that fewer goals also permitted improved metric measurability (U.S. Congress. Senate. Committee on the Budget. 2013). However, Moynihan (2013) argues broad agency goals and CAP goals tend to require collaboration with large stakeholder networks to develop and implement, which creates a complex environment of actor interests that make reaching a consensus on clear goals a challenge (12-13). It thus appears the effects of increased goal prioritization and increased stakeholder complexity to some degree offset each other, leading to no net gains in average performance goal clarity. This may help explain why federal managers had challenges developing meaningful measures.

There is mixed evidence concerning the transformation of agency cultures towards a results orientation with the values of employee accountability and empowerment in the past decade. Recalling the previous discussion of the values of the National Performance Review and

the purposes of GPRA which included ". . .improve federal program effectiveness and public accountability by promoting a new focus on results, service quality, and customer satisfaction,"[44] legislators and other stakeholders during the whole GPRA-PART-GPRAMA period advocated changing federal agency cultures to make them more results-oriented as a desired outcome in its own right, and certainly as a necessary condition for improved agency performance. Scholarly theorizing and empirical research support the assumptions of merit for public management practice in this aspiration.

Historical evidence during the GPRAMA period offers at least anecdotal examples of agencies with high and low levels of both accountability and manager empowerment. Interestingly, some empirical research suggests GPRAMA may have contributed to the creation of new incentive structures and business processes for federal and military leaders, enhancing accountability for their performance and/or empowered in their decision making. For example, Ames's (2015) study of the Coast Guard finds slowly improving accountability in its culture. He implicates the public relevance of Coast Guard operational functions, a bias towards action vice analysis, a suspicion of any proposal that would require greater procedural centralization, and promotion rewards for officers with an execution vice planning background for managers' lack of intrinsic motivation to define and use performance measures. Ames (2015) also finds senior leaders in the Coast Guard transitioning from using output measures like the number of tons of drugs seized to outcome measures like number of transnational crime networks disabled (131-132). Federal investigations have demonstrated mixed results concerning improvements in accountability and employee empowerment in agency cultures in the past decade. GAO surveys of federal managers (GS-13 through SES level) included questions on "agency climate" to

[44] Government Performance and Results Act of 1993, section 2(b)3.

collect federal manager perceptions on different indicators of culture. Six of these align thematically to the concepts of manager accountability and empowerment in federal agencies. Three questions showed minor improvement from 2013 to 2017, and three declined. Table 3.3 below compares the results for 2013 and 2017.

Table 3.3. Accountability and Empowerment Indicators in CFO Act Agencies: 2013 and 2017

Question from GAO Survey of Federal Managers	2013	2017	Net Change
1. Agency managers/supervisors at my level have the decision making authority they need to help the agency accomplish its strategic goals. (E)	35.8%	36.8%	1.0%
2. Agency supervisors/managers at my level are held accountable for agency accomplishment of its strategic goals. (A)	59.6%	54.2%	-5.4%
3. Agency supervisors/managers at my level are held accountable for the results of the program(s)/operation(s)/projects(s) they are responsible for. (A)	68.1%	64.6%	-3.5%
4. Employees in my agency receive positive recognition for helping the agency accomplish its strategic goals. (E)	41.9%	45.5%	3.6%
5. My agency's top leadership demonstrates a strong commitment to achieving results. (A)	59.7%	62.3%	2.6%
6. Agency managers/supervisors at my level use performance information to share effective program approaches with others.	34.3%	33.8%	-0.5%

Notes
1. Data shows percentage of managers responding "to a great extent" or "to a very great extent"
2. (A) accountability measure. (E) empowerment measure

Sources of data: U.S. Government Accountability Office. 2013. "Managing for Results: 2013 Federal Managers Survey on Organizational Performance and Management Issues," GAO-13-519SP (June 26, 2013). https://files.gao.gov/special.pubs/gao-13-519sp/results.htm#question_72; U.S. Government Accountability Office. 2017. "Supplemental Material for GAO-17-775: 2017 Survey of Federal Managers on Organizational Performance and Management Issues," GAO-18-776SP (Sep 29, 2017). https://files.gao.gov/special.pubs/gao-17-776sp/resultsall.htm

When the unit of analysis changes to all federal employees, and not just managers, the results are more encouraging and likely offer a less biased portrayal of cultural shifts than the GAO data

from federal managers taking a survey specifically on GPRAMA implementation, which could be narrowly understood by some managers to imply performance culture vice agency culture in general. Table 3.4 compares employee responses to questions related to accountability and empowerment in the 2015 and 2019 Federal Employee Viewpoint Surveys of the Office of Personnel Management (OPM):

Table 3.4. Indications of Accountable and Empowered Culture in the U.S. Government (2015 and 2019).

Question from Federal Employment Viewpoint Survey	2015	2019	Net Change
I am held accountable for achieving results (A)	81.0%	83.0%	2.0%
I know what is expected of me on the job (A)	79.0%	81.0%	2.0%
When needed I am willing to put in the extra effort to get the job done (A)	96.0%	96.0%	0.0%
Managers review and evaluate the organization's progress toward meetings its goals and objectives. (A)	59.0%	64.0%	5.0%
I feel encouraged to come up with new and better ways of doing things (E)	56.0%	62.0%	6.0%
I like the kind of work I do (E)	83.0%	83.0%	0.0%
Employees have a feeling of personal empowerment with respect to work processes. (E)	43.0%	50.0%	7.0%
Employees are recognized for providing high quality products or services. (E)	47.0%	54.0%	7.0%
Creativity and innovation are rewarded. (E)	37.0%	44.0%	7.0%
Managers support collaboration across work units to accomplish work objectives. (E)	54.0%	61.0%	7.0%

Notes
1. Data shows percentage of positive responses.
2. (A) accountability measure. (E) empowerment measure

Source of data: U.S. Office of Personnel Management (2019), *Federal Employment Viewpoint Survey: Governmentwide Management Report*, 28-31. https://www.opm.gov/fevs/reports/governmentwide-reports/governmentwide-management-report/governmentwide-report/2019/2019-governmentwide-management-report.pdf

Credible commitment of senior leaders played an important role in GPRAMA implementation, but the effects appear to have been concentrated at the departmental level. As discussed, the establishment of CAP Goal Leaders, COOs, PIOs, Deputy PIOs and Goal Leaders,

and Deputy Goal Leaders in each agency was designed to improve the reform responsiveness of these appointed and career officials. In theory, leaders who are more actively engaged in using performance information to manage programs, and include all goal contributors into performance reviews, will be perceived by subordinate managers and employees as more credibly committed to the reform.

Evidence points to some improvements in senior leader engagement in performance management during this period, especially in support of CAP goals, but, as with goal clarity and goal-setting routines, also superficial institutionalization inside many federal agencies. CAP Goal Leaders and teams reported CAP goal designation had the intended effect of greater attention and emphasis within their agencies on supporting the CAP goal for which the agency was the goal lead (U.S. Government Accountability Office 2015, 26). This suggests protecting their agency's reputation may have predictably been a motivator for senior leaders. Despite this, GAO found some goal contributors from other agencies, and even contributors within the agency, were not included in agency-level performance reviews by Goal Leaders[45] and that, government-wide, federal manager responses to questions about their agency's use of performance information were on average less than those of senior executive service (SES) managers (U.S. Government Accountability Office 2015, 21),[46] both serving as evidence of incomplete organizational institutionalization of performance review business processes.

[45] Participation of goal contributors from other agencies and from other parts of the agency was mandated by U.S. Office of Management Budget, *Delivering and Efficiency, Effective, and Accountable Government*, by Jacob J. Lew, OMB M-11-31 (Washington, DC, 2011). https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2011/m11-31.pdf

[46] GAO calculated the difference between non-SES and SES responses was statistically significant government-wide, and at the following agencies: DHS, VA, Treasury, Agriculture, Energy, HUD, NRC, USAID, and DOD (U.S. Government Accountability Office 2015, 57).

This claim is entirely consistent with the 2017 GAO survey of federal managers[47] which indicated that there was no statistically significant change since their previous survey in 2013 in managers' agreement to the statements "my agency's top leadership demonstrates a strong commitment to achieving results" and "my agency's top leadership shows commitment to using performance information in decision making," but there was a significant decline in the latter since 2007 (U.S. Government Accountability Office 2017, 42-43). Therefore, the preponderance of the evidence shows that, within federal agencies, the effects of credible leadership commitment, if any, are experienced mostly at the top of the organization. Based on perceptions of federal managers at all levels (GS-13 and higher), GPRAMA institutionalization has still not reached street level, or anywhere close to it, except in rare instances. This is not to suggest that street-level federal bureaucrats have no measures of performance, however. The author's interviews with six federal performance managers from five different agencies in March of 2021 indicated business units in their agencies maintain operational-level or program-level measures of performance, but these were often not integrated with GPRAMA-type measures at the departmental or agency level.

As discussed in chapter 1, employee training in performance measurement has been theorized as an important condition for institutionalizing performance management. At the start of the GPRAMA regime, employee training in strategic planning or performance assessment was recognized as deficient government wide. The Comptroller General testified that GAO analysis of a 2007 survey of federal managers had found only about half had received any; OMB and the PIC responded with a commitment in the President's FY 2012 budget to helping agencies

[47] The GAO invited a stratified random sample of 4,395 managers out of a population of 153,777 mid-level and upper-level agency managers and supervisors at 24 CFO Act agencies and received a 67 percent response rate (U.S. Government Accountability Office 2017, 4).

improve employee skills in analyzing and using performance information (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 65). Furthermore, to ensure consistency and standardization in developing employee skills in performance management, GPRAMA required the Office of Personnel Management to define and catalog performance management skills and competencies (U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs 2011, 65). Empirical research offered evidence that federal employee training in performance management was significantly and positively related with the use of performance goals, but not of performance information, and that the training that was offered explained the reform to employees, but offered little in terms of building workforce capacity to execute performance management tasks (Kroll and Moynihan 2015, 416-417). GAO statistical modeling of its 2017 survey of federal managers indicates out of five possible impacts of performance management training,[48] only one was perceived by federal managers in 2017 to have a statistically significant relationship with training: training helped managers ". . .to link program(s)/operation(s)/project(s) to the achievement of agency strategic goals" (U.S. Government Accountability Office 2018, 95).

Capacity for evaluation continued to be an important condition for agency implementation of GPRAMA. Limitations on management attention which led to use of priority goals and new performance management roles have already been discussed. Because PART had required agencies to tap their own resources to evaluate and report their performance, in 2009, OMB gave guidance to agencies to request funding in FY 2011 for additional capacity to conduct in-depth program evaluations, and noted the Obama Administration had appointed

[48] The other four were manager "use of performance information to make decisions," "set program performance goals," "develop program performance measures," and "assess the quality of performance data" (U.S. Government Accountability Office 2018, 95).

Cabinet and sub-Cabinet officers who had extensive experience with performance measurements to improve government performance leveraging effective practices from state and local governments (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 11). Within agencies, the need for cross-functional integration between strategic management functional communities responsible for performance management, budgeting, and policy was identified as a necessary improvement for the new performance management framework (U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee 2009, 28). As understood in the inferential model, such integration between silos of agency expertise makes better use of existing evaluative resources, and so represents an example of capacity building at no additional financial cost.

During the GPRAMA period, empirical evidence for the effects of agency identity on GPRAMA institutionalization within federal agencies became more available with increased GAO analysis and reporting, which GPRAMA required for compliance monitoring. GAO investigative findings affirmed the predictions of bureaucratic behavior associated with James Q. Wilson's taxonomy discussed in chapter 1. By 2015, OMB and federal agencies had not yet implemented 55 of 69 GAO recommendations (80 percent) since 2011 to improve their implementation of GPRAMA (U.S. Government Accountability Office 2015, 76). Analyzing the distribution of 28 GAO recommendation non-implementations by federal agencies (excluding OMB and OPM as their GPRAMA activities relate to external support for other agencies) in three compliance practice categories, and categorized by Wilson bureaucratic type, a striking

pattern emerges. Here, descriptive analysis alone is enough to be persuasive. Table 3.5 presents this data.

Table 3.5. Government Accountability Office Recommendations Not Implemented by Federal Agencies (2015).

Bureaucratic Type/Agency	**Category of Non-Implemented GAO Recommendation**			
	Quality of Performance Information (1)	**Data-Driven Reviews (2)**	**Customer Service Performance (3)**	**TOTAL (4)**
Production Agencies				
Commerce				-
GSA				-
Labor	2			2
NRC				-
NSF				-
SBA				-
SSA				-
USAID				-
VA			1	1
Sub-Total (% Total)	2 (18%)	0 (0%)	1 (20%)	3 (11%)
Average Per Agency	0.22	0.00	0.11	0.33
Procedural Agencies				
Education			1	1
HUD				-
OPM				-
Sub-Total (% Total)	0 (0%)	0 (0%)	1 (20%)	1 (4%)
Average Per Agency	0.00	0.00	0.33	0.33
Craft Agencies				
Energy				-
EPA				-
NASA	2			2
Sub-Total (% Total)	2 (18)%	0 (0%)	0 (0%)	2 (7%)
Average Per Agency	0.67	0.00	0.00	0.67
Coping Agencies				
Agriculture	2	4	1	7
Defense	2	3		5
DHS	1	1	1	3
HHS		1		1
Interior	2		1	3
Justice				-
State		3		3
Transportation				-

Treasury				-
Sub-Total (% Total)	7 (64%)	12 (100%)	3 (60%)	22 (79%)
Average Per Agency	0.78	1.33	0.33	2.44
Total all Categories	11	12	5	28

Notes:
1. Improve Public Reporting on Quality of Performance Information for Agency Priority Goals
2. Strengthen Data-Driven Review Practices
3. Integrate Performance Goals, Standards, Targets into Customer Service Efforts
4. Agencies having no GAO recommendations, or which implemented all GAO recommendations, are shown with a "-"

Source of data: U.S. Government Accountability Office. 2015. *Managing for Results: Implementation of GPRA Modernization Act has Yielded Mixed Progress in Addressing Pressing Governance Challenges*, GAO-15-819 (Washington, DC, 2015), 97-102, accessed April 22, 2021, https://www.gao.gov/assets/gao-15-819.pdf

As a group, Coping bureaucracies accounted for 78.6 percent of these, and these agencies had an average number of non-implementations per agency (2.44) that was 3.6 times higher than the next lower category of Craft bureaucracies (.67), which was, in turn, double that of Procedural and Production bureaucracies (both .33). The ways in which the agencies experienced difficulties with implementing the GAO recommendations also underscore the relationship between their agency identity and their non-implementation. Coping agencies, which have difficulty measuring both outputs and outcomes of their programs, accounted for 63.6 percent of non-implementation of GAO recommendations to improve public reporting on quality of performance information that is used to measure progress towards achieving agency priority goals. Craft and Production agencies accounted for 18.2 percent of these non-implementations each (total of 4) but these were concentrated in only two agencies: Department of Labor and NASA. As a Craft agency, NASA should be able to measure observable outcomes but have difficulty with measuring less-observable outputs. As a Production agency, Labor should theoretically have little difficulty measuring outcomes or outputs and so should be well

situated to implement this GAO recommendation; however, it may also be the exception that proves the rule, as it is the Production agency whose managers' combined reported use of output and outcome measures in the 2017 GAO survey of federal managers was the lowest of all agencies in the Production category, followed closely by the Veterans Administration. The latter was, interestingly, the only Production agency to fail to implement GAO recommendations to integrate performance goals, standards, and targets into customer service efforts.

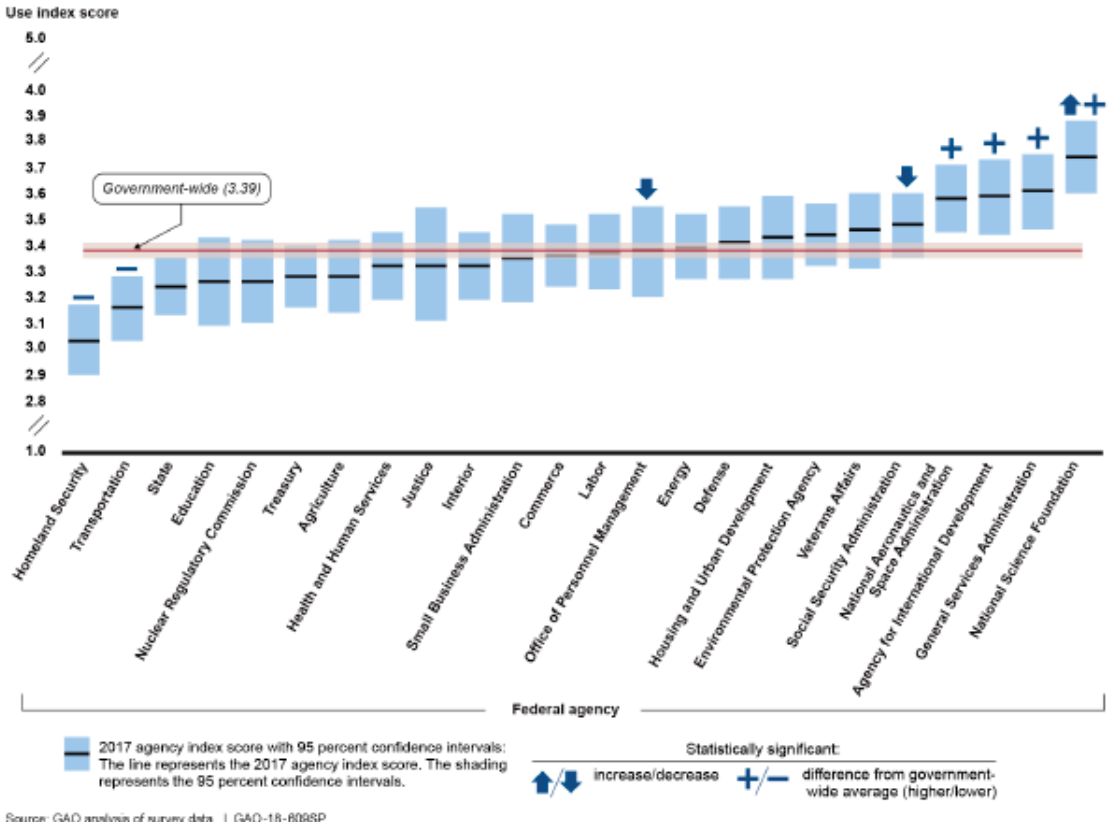
Coping agencies had a near "monopoly" on non-implementation of GAO recommendations to strengthen practices for data-driven reviews, with the Department of Agriculture at the top of the agency pile with four and the Departments of Defense and State with three each. GAO found these agencies were deficient in terms of the agency head or COO not running review meetings, involving Goal Leaders and officials with functional management responsibilities, or scheduling review meetings at least quarterly. Defense failed to implement additional GAO recommendations to ". . .discuss at-risk goals and improvement strategies. . ." ". . .assess whether specific program activities, policies, and other activities are contributing to goals as planned. . ." and to determine and track follow-up actions (U.S. Government Accountability Office 2015, 99). The Department of Defense may be a unique case. Interviews with several Deputy PIOs by the author in the spring of 2021 revealed that many PIC members believe OMB gives the Defense Department "a pass" on GPRAMA requirements because of its size, complexity, and its preponderant responsibilities in implementing the National Security Strategy.[49] No such exception was identified during Deputy PIO interviews regarding the Department of Agriculture.

[49] The Department of Defense (DoD) *Fiscal Year 2021 Annual Performance Plan* (APP) (U.S. Department of Defense 2020), for example, has very little to do with outcome-based performance as required by GPRAMA, except in the limited sense that military readiness may be

As noted previously, the 2017 GAO agency survey of federal managers found no statistically significant difference in their use of performance information government-wide from 2013 to 2017. It did, however, find six federal agencies had a statistically significant variance from the federal average score for perceived use of performance information, a key metric of institutionalization in this study, in 2017. Figure 3.2 below is from the GAO report (U.S. Government Accountability Office 2018, 15). It shows that the manager-reported average use of performance information in the DHS and DOT was significantly less than the federal average, and the manager-reported average use at NASA, USAID, GSA, and NSF were significantly above the federal mean.

Figure 3.2. Government Accountability Office Estimates of Average Manager Perceptions of Performance Information Use, by Federal Agency (2017)

considered an outcome which is identified, but not measured (or perhaps even measurable). The APP is overwhelmingly focused on improving internal mission support, business operations and management functions such as logistics and supply chain management, financial management, contract management, human capital, business systems, IT, etc. While these help DoD to generate military power, they are at best examples of outputs and mostly just internal processes. Instead of a Chief Operating Officer (who should be the Deputy Secretary of Defense as per GPRAMA requirements), the document was signed by the DoD Chief Management Officer (CMO), a subordinate appointed senior executive. The APP itself is consigned to appendix one of the annual document, with the main document being the "National Defense Enterprise Business Operations Plan." This reveals how DoD has chosen to implement GPRAMA in a way that supports internal management reform needs, not outcome-based performance accountability as per legislative intent. It appears they have been successful in achieving a waiver by estoppel.



Source: U.S. Government Accountability Office. 2018. "Managing for Results: Government-Wide Actions Needed to Improve Agencies' Use of Performance Information in Decision Making." Government Accountability Office GAO-18-609SP. Washington, DC: Government Accountability Office. https://www.gao.gov/assets/700/694269.pdf.

The pattern is startling when seen through the lens of Wilson's taxonomy. As discussed in chapter 5, this study classifies DHS and DOT as Coping agencies, NASA as a Craft agency, and USAID, GSA, and NSF as Production agencies. At least to the extent these GAO data are significant, Coping agencies are behind other types of agencies in institutionalizing the use of performance information. The pattern of strong use of performance information by Production agencies seems clear enough to justify the use of the more advanced statistical methods this study will apply to test hypothesis H_{9b}, "a significant relationship exists between an agency's identity and its use of performance information" in chapter 5.

Current Developments and the Way Ahead

With the advent of the Biden Administration in January 2021, as it has with all previous administrations, change in the governance of federal performance management is once again in the air. It is, however, still too soon to predict specific changes in the Federal Performance Management Framework. On December 23, 2020, in the twilight of the Trump Administration, OMB issued a memo to federal agencies that unexpectedly (according to at least two deputy PIOs the author interviewed in March, 2021) rescinded part 6 of OMB Circular A-11, which had been the core guidance for all federal agencies on GPRAMA implementation, without relieving them of their statutory obligations. The OMB director justified his decision with the arguments that performance information produced under GPRAMA was little used and that "GPRA. . . has yielded few tangible results. Indeed, the thousands of pages of performance data generated by agencies and posted on performance.gov each year attract little interest."[50] The memo indicated additional guidance would be forthcoming. OMB memo M-21-13 to federal departments and agencies, dated January 15, 2021, was the first installment of fresh guidance. It indicated a goal to revisit how performance management laws should be implemented and presented a bleak assessment on the cost- effectiveness of GPRAMA to date:

> There is scant evidence of benefit for the time and expense that agencies annually invest in compliance with these laws under the legacy practices that have been employed to date. Congress and the public have long turned a blind eye to the performance information produced, and senior Executive Branch officials struggle to meaningfully incorporate the information into management efforts due to the questionable quality of the data yielded. This situation should not go unaddressed; valid, meaningful performance data is critical to oversight and adaptation of Executive Branch operations.[51]

---

[50] U.S. Office of Management and Budget, *Revisions to OMB Circular No A-11 dated July 10 2020* by Russell T. Vought (Washington, DC, 2020), https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf

[51] U.S. Office of Management and Budget, *Implementation of Performance Management Statutes* by Russell T. Vought (Washington, DC, 2021), https://www.whitehouse.gov/wp-content/uploads/2021/01/M-21-13.pdf

OMB M-21-13 also acknowledged the government had failed to implement the GPRAMA requirement to post an inventory of all federal programs and their associated costs on the public website Performance.gov for stakeholder use to compare and contrast programs due to continuing agency differences over how to define a "program" and incompatible program databases, then announced it was leading an exploratory pilot to better understand the sources of program data incompatibility.

Summarizing the history of GPRAMA institutionalization, perhaps Kamensky (2021) expresses it best: ". . .it institutionalized a governance framework, a performance management framework, and incorporated a set of new administrative routines that have helped institutionalize the use of performance management in agencies, *at least at the highest levels*. A key remaining challenge is to incorporate these routines at lower levels within each agency" (20, italics added). The evidence and analysis in this chapter demonstrates that the variance in GPRAMA institutionalization between federal agencies is every bit as pronounced as it is between organizational levels within single agencies. Based on the evidence available from the modern history of performance management reform in the U.S. Government, there are sufficient grounds to postulate that the conditions of external political influence and resource dependency, credible senior agency leadership commitment, empowered and accountable agency culture, and agency identity (here defined by James Q. Wilson's 1989 taxonomy of bureaucratic types) contributed to between-agency variance in GPRAMA institutionalization, either by their presence or, in some cases, their absence. These conditions and their causal mechanisms require more research than presently exists in the literature to fully understand how they, along with the conditions of oversight, goal clarity, training, capacity for evaluation, and measurement system maturity, combine to lead U.S. federal agencies to institutionalize GPRAMA.

This historical analysis provided useful insights into these conditions during the GPRA, PART, and GPRAMA eras, and offers clues as to how they may interact with each other. Historical evidence alone, however, is insufficient to prove their effects on institutionalization in federal agencies. The remainder of this dissertation explores the effects and relationships of the system and organization action situation conditions more fully, extends the historical analysis using mixed methods of research design using primary data sources from 2017 and 2021 and, based on the findings, proposes answers to this study's research questions.

Chapter 4. How System and Organizational Conditions Affect GPRAMA Institutionalization

The literature review in chapter 2 presented a framework for analysis building on organizational and public management theories to guide the examination of sources of variance in agency institutionalization processes, including practices related to performance management reform. Primary theoretical perspectives discussed included categories of neo-institutionalism, organizational identity, and the phases of institutionalization of new practices. Chapter 2 also analyzed and synthesized a significant corpus of public management theory and empirical research on system action and organization action conditions that leading scholars have identified as probable significant and generally positive influences on public agency institutionalization of performance management reforms. For each condition, it also proposed hypotheses derived from the literature to test, and introduced how the study modeled it in the inferential framework and operationalized it for measurement. Chapter 3 employed a historical institutionalist lens to characterize the evolution of the institutionalization of GPRA, PART, and GPRAMA in the U.S. Government. The history showed systemic and organizational conditions exerted varying but significant levels of influence on reform institutionalization within federal agencies and government-wide over the nearly three decades since GPRA was promulgated in 1993.

In this study, the final stage of the institutionalization process, sedimentation--the point at which practices become deeply embedded in routines and are considered as a given in daily work life by members of the agency--is of primary interest because it offers the best opportunity to understand which system action and organizational conditions best explain the final outcome of incomplete GPRAMA institutionalization, i.e. institutionalization that stops at the objectification

or habitualization stages. Understanding the relative strength and salience of conditions, during a period when the maximum feasible institutionalization should reasonably have been attained given the governance context and timeline (2017-2021, 24-28 years after GPRA was signed), and their combined effects on institutionalization outcomes, is the key to unlocking insights about the practical limits of accountability-based performance management reform in the U.S. Government.

The purpose of chapter four is to answer the first research question. It proceeds by first describing the mixed methods research strategy. For the quantitative and then the qualitative research, it presents the research design, the results and findings, and then interprets the findings to draw modestly generalizable conclusions about the influence of the system action and organizational conditions, and their configurations, that influence the institutionalization of GPRAMA in the U.S. Government and in individual agencies. While this study models agency identity as a condition in the organization action situation, the discussion of the research design, results, findings, and interpretation related to agency identity and bureaucratic type (a taxonomy of agency identity) is reserved for chapter 5 in order to provide a focused analysis and response to the second research question

Research Strategy

The research purpose and the nature of the research questions and hypotheses suggest a mixed methods research strategy would be effective for studying the effects of system action situation and organization action conditions. Hicklin (2010) notes combining quantitative and qualitative research methods can make scholarship on the management of government agencies more complete by permitting the researcher to investigate research subjects (in this case, agencies and managers) in depth while still enabling generalizability of research results from

large samples (265). Hendren, Luo, and Pandey (2018) claim mixed methods ". . .add value by corroborating evidence, illuminating perspectives, and balancing the biases inherent in using qualitative or quantitative methods alone" (904). Ragin (2014) affirms the value of a combined quantitative and case study approach: "ideally, a combined strategy should allow the investigator to consider both structural factors and factors reflecting historical processes and human agency. In short, one strategy should check the biases of the other" (70). The multiplicity of predictive conditions and dependent variables in this study, and the complex and interdependent nature of these conditions, call for a mixed methods research design that not only identifies the most influential conditions and causal combinations using quantitative models, but also generates evidence concerning configurational causal pathways within federal agencies based on comparative qualitative data collection and analysis.

Campbell and Stanley (1967) suggest a correlational analysis can disconfirm low-credibility hypotheses: ". . . the absence of such a correlation can rule out many simple, general, causal hypotheses, hypotheses as to the main effect of X. In this sense, the relatively inexpensive correlational approach can provide a preliminary survey of hypotheses, and those which survive this can then be checked through the more expensive experimental manipulation" (64). Following an approach proposed by Amenta and Poulsen (1994), endorsed by leading QCA methodologists (e.g. Yamasaki and Rihoux 2009), and employed by scholars of public policy (e.g. Fischer, Kaiser, and Rohlfing 2006), this design applies Campbell and Stanley's insight and first uses a quantitative design employing cross-sectional analysis of random and fixed effects regression models to analyze a large-N dataset of federal manager perspectives on a wide range of matters concerning GPRAMA implementation created in 2017 in order to confirm which hypothesized independent variables from the literature demonstrate statistically-significant

correlations with the dependent variables (here, the two GPRAMA institutionalization indicators), and the relative strength of the associations. It then selects those predictors which prove to be both significant and substantial as conditions for a quasi-experimental qualitative exploratory investigation using systematic comparative case analysis employing fuzzy set QCA techniques.

Among other purposes, QCA is useful for summarizing research data, checking its coherence, and developing new theoretical claims (Berg-Schlosser et al. 2009, 15). The goals of the qualitative investigation using QCA techniques are thus to describe, order, and make sense of noisy data from a small-N survey of federal performance managers in 2021, and then to identify the most likely causal configurations that, based on the presence or absence of the selected conditions and with some level of consistency, lead to the institutionalization of GPRAMA practices in federal agencies. The findings on causal pathways generate original empirically-derived theoretical arguments about the necessity and sufficiency of causes and their configurational patterns. Finally, descriptive comparative case analysis of federal agencies based on survey and interview data with federal performance managers illustrates empirical micro-level contexts and explains how organizational conditions affect agency institutionalization of GPRAMA in ways that cannot be derived from quantitative or QCA-enabled research. Table 4.1 presents the relationship of the data sources and mixed methods with elements of the research questions in this study.

Table 4.1. Research Question Elements, Analytic Strategy, and Data Sources

RQ	**Research Question Elements**		**Analytic Method**		
			Quantitative	**Qualitative**	
#	**Predictor Variable**	**GPRAMA Outcome**	**HLM**	**fsQCA**	**Case Study**
1	Political-Resources	PM Adoption		S	S,I
1	Political-Resources	Use of PI		S	S,I

1	External Oversight	PM Adoption	G		
1	External Oversight	Use of PI	G		
1	Goal Clarity	PM Adoption	G		
1	Goal Clarity	Use of PI	G		
1	Accountable Culture	PM Adoption	G	S	S,I
1	Accountable Culture	Use of PI	G	S	S,I
1	Leadership Commitment	PM Adoption	G		
1	Leadership Commitment	Use of PI	G	S	S,I
1	PM Training	PM Adoption	G		
1	PM Training	Use of PI	G		
1	Evaluation Capacity	PM Adoption	G	S	S,I
1	Evaluation Capacity	Use of PI	G	S	S,I
1	Measurement Maturity	PM Adoption	G	S	S,I
1	Measurement Maturity	Use of PI	G		
2	Agency Identity	PM Adoption	G		
2	Agency Identity	Use of PI	G		
2	Bureaucratic Type	PM Adoption	G		
2	Bureaucratic Type	Use of PI	G		
1,2	Manager Incentives	PM Adoption			I
1,2	Manager Incentives	Use of PI			I
1,2	Manager Bias/Attitude	PM Adoption			I
1,2	Manager Bias/Attitude	Use of PI			I

Notes:
RQ Research Question
HLM Hierarchical Linear Model
fsQCA Fuzzy Set Qualitative Comparative Analysis
Data Sources (G = GAO Survey, S=2021 Survey, I=Interviews)
Shaded regions indicate intentionally omitted based on quantitative findings

The chapter proceeds with a detailed discussion on the quantitative and qualitative research designs. Each design discussion will cover its data sampling strategy, data collection and preparation procedures, the construction of variables (quantitative) or conditions (qualitative), and units of analysis and measurement models using the operationalizations already described briefly in chapter 2. In the conclusion of the chapter, I present and integrate the quantitative and qualitative findings to postulate causal configurations of conditions that lead to positive and negative GPRAMA institutionalization outcomes. Assumptions and limitations of the research are presented in chapter 6.

Quantitative Research Design

Introduction

The goal for the quantitative research in this study is to analyze the system action situation and organization action situation conditions, here modeled as predictor variables, and the degree of their association with the two dependent variables which are proxies for measuring GPRAMA institutionalization: the adoption of performance measures and the use of performance information. A related objective is to empirically validate or falsify claims in the literature about the relative importance of each condition to agency institutionalization decisions in the context of the U.S. Government. The theoretical expectation is that some conditions exert greater influence on incentive structures facing federal managers than others do, and so deserve greater attention and emphasis in research by scholars and practitioners of public management. This chapter will consider all conditions in the inferential model (figure 2.1) except agency identity/bureaucratic type, which are modeled separately in chapter 5 to better focus analysis and findings to answer research question two.

Data Source and Sample

As with related research by Dull (2009), Kroll and Moynihan (2015), and Moynihan and Kroll (2016), this study takes advantage of GAO investments in government-wide large-N surveys, which would be far beyond the means and access of mortal scholars to orchestrate, to obtain cross-sectional data from a large sample of U.S. Government managers. The author is grateful to the GAO for sharing its full dataset from its November 2016 to March 2017 online survey of Organizational Performance and Management Issues with a nationwide sample of 4,395 mid- to senior-level managers (U.S. Government Accountability Office 2017b). The GAO had previously administered substantially the same survey in 2000, 2007, and 2013, although the

survey series is not in a panel format. Surveyed managers were career officials who ranged from grade GS-13 to Senior Executive Service and were randomly selected out of a population of 153,779 mid- and senior-level managers across the 24 federal agencies subject to the Chief Financial Officers Act of 1990. GAO selected managers at random from the Office of Personnel Management Enterprise Human Resources Integration database as of September 15, 2015 (U.S. Government Accountability Office 2017a, 4), therefore systemic self-selection bias is unlikely. The GAO structured the dataset to make it generalizable to the population of all U.S. federal managers using agency-level sample weights to account for agency manager population sizes; there were 2,726 responses, a 67 percent response rate (U.S. Government Accountability Office 2017, 4). For each observation (i.e., the full set of individual manager responses), the data identified the federal agency, permitting agencies to be units of observation via clustering of observations of manager responses.

It is important to note that the survey data measure manager perceptions about their organizations, the external environment, and their subjective experience with whatever incentives shape their responses to daily work and the survey vehicle itself. In turn, all of these are mediated by the system action and organizational conditions postulated in the inferential model shown in figure 2.1. As a result, the findings are mathematical abstractions that measure co-variance between different sets of manager perspectives on GAO questions related to specific conditions, not objectively observable or recorded actions or behaviors, or even manager perceptions on relationships between the conditions (the GAO did not ask them about these).

Measuring manager perceptions as a basis for drawing inferences about the force of institutional pressures and incentives on manager decisions should be a method with high construct validity, at least according to sense-making theory, which holds that people make sense

of their environment and develop meaning through interactions with the environment and other people (Powell and Rerup 2017; Weick 1993). Institutions and organizations have no meaning except in the minds of individuals subject to them, and there is no other way to collect data on their effects on sense-making without tapping into people's subjective experiences with them. No archival or personal data on individual managers is available in the GAO dataset due to privacy rules, so personal data about age, education level, gender, race, sex and other personal factors, commonly researched in the performance management literature (Kroll 2015b), are not available for use as control variables in this study. An alternative and unconventional approach for control variables is presented in the section on measurement models.

Data Collection and Preparation Procedures

As noted, GAO collected the data as part of its 2017 survey of federal managers for the quantitative portion of this study. A critical task of this study, however, was the careful and appropriate selection of which survey questions would be incorporated into the factor analysis that was used to develop the latent variables used in the in models of this study, described in the next section. Survey questions (appendix C) were considered if they were theoretically aligned to the conceptualization of the conditions in the system action and organizational action situations as described in the literature review. This was an analytic decision that was admittedly subjective but informed by the author's extensive review of the performance management literature and his three decades of service as a manager in the federal government. The selection of questions was also presented to and reviewed by fellow PhD students and faculty members in the Virginia Tech Center for Public Administration and Policy in several academic forums.

Not all possible theoretical indicia in the study have a suitable proxy in the GAO survey dataset (and the GAO survey asked many questions not thematically aligned to study indicia), so

the variable operationalization and quantitative analysis is necessarily constrained by the availability of suitable questions in the GAO survey. For example, the dataset did not contain any suitable proxies with which to validly operationalize the influence of politics and resource dependency on federal manager perspectives, so these are measured using qualitative methods described in the section on the qualitative research design to follow. Appendix C shows the GAO survey questions which are theoretically-aligned to system action and organization action situation conditions, which are treated here as the dependent and independent variables for the quantitative models, and how each question used is aligned to each variable, along with the definitions of the latent variable. No question was used for more than one model variable. Appendix C also presents descriptive statistics about government-wide federal manager responses to each survey question as well as the factor variables with which they were associated after their estimation using exploratory factor analysis.

I prepared and cleaned the GAO dataset to support the regression modeling analysis. First, because most of the GAO survey data was structured using a Likert scale from 1 ("Very Great Extent") to 5 ("No Extent") and 6 ("No Basis to Judge"), I reverse-coded the survey variables by subtracting them all from 6 to ensure higher scores would reflect a more positive manager evaluation of the question, because scaling up to more positive responses is more intuitive and aids with interpretability.[52] As a result, "No Extent" responses thus became 1 and "Very Great Extent" became 5. "No Basis to Judge/Not applicable" responses became zero and were excluded from the subsequent analysis. I did this for all Likert-scale variables used in the

[52] Series 8 questions were already coded in the desired manner based on the logic of the question, so no change was made to any of these.

model. Dichotomous variables in series 13 related to performance management training were likewise reversed to make 1 "yes" and 0 "no" for the same reason.

On first evaluation, I found there was a substantial amount of seemingly missing data in the dataset fields of interest. After further review, I found much of it was "planned" missing due to survey instructions to skip questions under certain conditions, and thus should not be formally considered as missing per se, while other missing data was missing for no clear reason. To deal with this, I recoded missing data into different categories of "missingness." I replaced missing data with Stata-standard special designators for different types of missing data depending on whether they were missing at random or completely at random (.) or "planned missing" (.#). For example, the first row in table 4.2 below shows the approach I used to deal with guidance after question 4 for respondents to skip all series 5 questions if their programs did not have performance measures, or if they did not know whether or not they did. Likewise, the GAO survey instructions directed managers at the Nuclear Regulatory Commission (NRC) to skip series 17 questions because the question did not apply to their agency.

Use of different missingness designators allowed me to calculate the rate of data missingness in a way that included data missing at random (MAR) or completely at random (MCAR) (the "." flag), but excluded "planned" missing data, or missing not at random (the ".a" and ".b" flags). "Not applicable/No basis to judge" responses, which are not considered "planned missing," initially recalculated as a zero via the reverse-coding procedure, were recoded again as a separate category of "hard" missing data (.c) to ensure that observations with this value would also be ignored by Stata during model fitting.[53] Table 4.2 shows the

[53] Guidance from my dissertation committee and additional reading of articles that used survey data for similar research goals indicated it is best to exclude Likert scale responses "not applicable" or "do not know" when fitting regression models. This effectively reduces the

replacement procedures to address the problem of missing data. Table 4.3 shows the calculated rate of missing data for survey questions with a missing rate above 3 percent after recoding. The average rate of truly missing data for all questions used in this study was 2.4 percent, and planned data and "Do not know/Not applicable" responses combined constituted 83.9 percent of all data that were skipped in modeling. While fitting the regression models, Stata skips observations with data coded as missing using listwise deletion.

Table 4.2. Codes for Categories of "Missing" Data in the 2017 GAO Dataset

Survey Questions	Recode	Reason	Condition
Series 5 questions	.a	GAO-directed skip	if question 4>1
Series 17 questions	.a	GAO-directed skip	if q16=3
Series 17 questions	.b	GAO-directed skip	if agency=17 (NRC)
Series 8 questions (reverse order kept)	.c	"Not applicable/Don't know" responses	If [question value]=6
All questions except series 8, and series 13	.c	"Not applicable/don't know" responses	if [question value]=0

Table 4.3. Analysis of Missing and Skipped Data in the 2017 GAO Dataset[54]

Question	Missing	Planned Missing or NA/Don't Know	Observations with Data	Missing Rate
17c	174	898	1,679	6.3%
17e	167	636	1,948	6.1%
17b	164	695	1,892	6.0%
17d	164	643	1,944	6.0%
11b	109	1,030	1,612	4.0%

Construction of Variables

sample to surveyed federal managers with a degree of knowledge and awareness of the situation in their organization related to the question and to those for whom the question is applicable.

[54] Only GAO survey questions with a missing data rate greater than 3 percent are shown. Overall mean of missing rates for all questions was 2.4 percent.

Appendix C defines the dependent and independent variables used in all regression models in this study and provides summary descriptive statistics. Each was conceptualized based on an interpretation of the performance management literature and operationalized in chapter 2. Each construct is developed as a latent variable derived from factor analysis that measures the common variance of the manager responses to the GAO survey questions associated with it, as shown in appendix C. Appendix E shows descriptive data about the latent variables used in models 1 and 2, discussed later, by federal agency.[55] The table at appendix E shows, for each CFO Act agency, the number of observations and mean value. The observations of individual federal manager responses are aggregated by agency based on an agency stratification field in the GAO dataset. All variables are standardized to facilitate comparisons, so the mean is the number of standard deviations from the universal mean of all 24 agencies. The number of observations is neither proportional to, nor reflective of, the number of managers in each agency. To normalize the data for analysis, the GAO dataset also includes a sample weight for each observation that accounts for cross-agency variance in the number of GS-13 to SES managers. This study combines responses from all managers, regardless of GS grade.

The use of continuous factor variables that integrate multiple theoretically-related indicators into estimates of latent constructs permit more reliable and valid analysis to test theoretical propositions in the literature than simply performing dozens of separate ordered logistics regressions between individual survey questions within the same construct, analyzing

[55] Note that the number of observations for PI is consistently less than it is for PM for all agencies. This is due to two reasons: 1) the inclusion of an indicator for management use of performance information for contracting decisions in the PI factor variable with fewer responses due to relatively fewer managers being aware of contracting activity in general much less how PI is used in it, and 2) the aggregation of missing data in the factor variable from the intersection of all truly missing data across 9 indicators, vice 5 for PM. As explained later, the analysis mitigates most of this discrepancy through the use of multiple imputation techniques.

them separately, and then trying to make sense of them to draw high-level inferences from the individual results. Rather than simply treat the ordinal data "as if" they were continuous, which would erroneously assume a constant relationship between ordered values, models are estimated using a polychoric correlation matrix as the input to an exploratory factor analysis using an approach proposed by Lee, Poon, and Bentler (1995).[56] According to Kolenikov (2016), "polychoric correlation is the correlation between two ordinal variables obtained as the maximum likelihood estimate under the assumption that the ordinal variables are obtained by coarsening a bivariate normal distribution."[57] In other words, the technique permits the estimation of latent continuous variables that would be the basis for ordered categorical variables if the categorical variables classified the continuous data by ordinal category, such as those categories used in Likert-scale surveys. To implement this in the study, I employed a supplemental module for Stata developed by Kolenikov (Kolenikov, n.d.). Comparing the polychoric correlation matrix it generates for GAO survey series 5 questions with a standard correlation matrix in which ordered data are treated "as if" they were continuous data in figure 4.1, the pairwise correlations between survey variables using polychoric correlation are greater than those for ordinary correlations, suggesting the polychoric method is able to associate more of the co-variation between two sets of ordinal data than treating ordinal scale data as continuous and applying Pearson regressions to it.

[56] For more information on the technique, see https://stats.idre.ucla.edu/stata/faq/how-can-i-perform-a-factor-analysis-with-categorical-or-categorical-and-continuous-variables/
[57] https://ideas.repec.org/p/boc/scon16/15.html

Figure 4.1. Comparison of Pairwise Correlation Matrixes: Polychoric and Pearson Models

Polychoric Pairwise Correlation Matrix

	E5a	E5b	E5c	E5d	E5e
E5a	1				
E5b	.60906632	1			
E5c	.4429735	.62614742	1		
E5d	.51215045	.66305422	.7415078	1	
E5e	.59880265	.6336405	.59357606	.69451484	1

Standard Pairwise Correlation Matrix

	E5a	E5b	E5c	E5d	E5e
E5a	1.0000				
E5b	0.5267	1.0000			
E5c	0.3725	0.5593	1.0000		
E5d	0.4335	0.5978	0.6836	1.0000	
E5e	0.5105	0.5614	0.5288	0.6238	1.0000

For each dependent and independent variable described in appendix C, I generated polychoric correlation matrices and then used the Stata FACTORMAT command to generate the exploratory factor analysis. Appendix C shows the Eigenvalues, factor loadings, average inter-item correlation (AIC), and Cronbach's alpha, a measure of internal reliability (consistency) that shows how inter-related items in a group are. All the constructs yield values exceeding generally accepted standards. Appendix D shows descriptive information on the 2017 GAO survey question data.

For each latent variable, factor analysis identified a single factor with Eigenvalues of greater than 1, ranging from 2.496 for agency capacity for performance evaluation to 6.836 for use of performance information. Therefore, no rotations were needed or performed. All factor

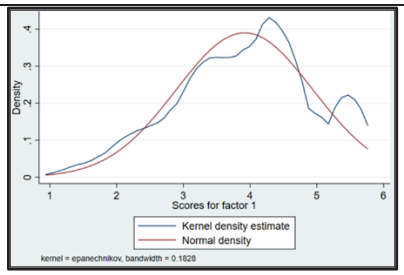
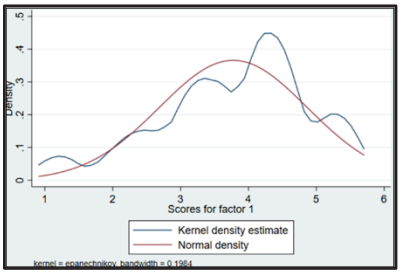
loadings are acceptable.[58] Alpha values for all latent variables are above .80 but only one, the dependent variable for manager use of performance information (PI), is above .95 (.952). Nine scale variables are included in PI, and all are conceptually similar with the highest inter-item covariance of any model variable (.686). But each scale variable in PI is a substantively important example of managerial use of PI, and each is important theoretically for the study, so I accept the high alpha value for PI in order to retain all the scale values in the latent construct to support subsequent modeling and analysis. The Kaiser-Meyer-Olkin procedure was employed to test sampling adequacy. The results were all above 0.50, the lowest threshold for acceptability recommended by IBM, with all but five survey questions above .70, confirming the underlying influence of the latent variables on the variance of the survey variables.[59]

I then conducted tests for normality of distributions and considered the need for possible transformations on all latent variables. Figure 4.2 shows results for the kernel density test for each. The graphs, which resemble a smoothed histogram, are near normal, but exhibit a characteristic "bumpiness" one would expect of a variable distribution derived via polychoric correlation using a set of ordered variables with five categories (or just two in the case of the dichotomous training variable, T). The second dependent variable, PI, had a similar kernel density distribution as shown in figure 4.2, although the density is, predictably, relatively higher near the ordinal thresholds of the GAO data.

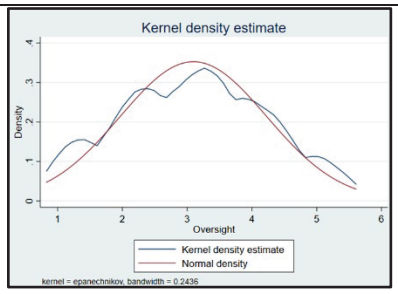
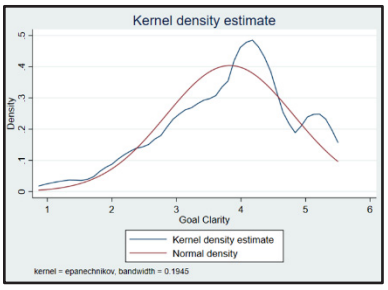
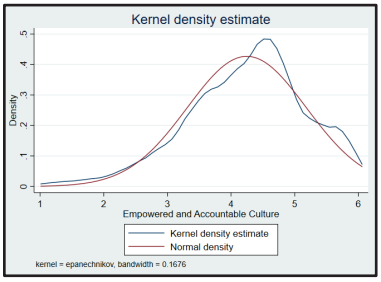
Figure 4.2. Distribution of Latent Variables Versus Normal Distribution

Dependent Variables	**Kernal Density Estimate**

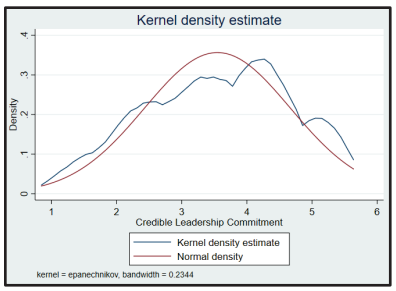
[58] Tabachnik and Fidell (2007) cite Comrey and Lee's (1992) recommendations for the following loading level interpretation: 0.32 (poor), 0.45 (fair), 0.55 (good), 0.63 (very good) or 0.71 (excellent) (649).
[59] https://www.ibm.com/docs/en/spss-statistics/23.0.0?topic=detection-kmo-bartletts-test.

Adoption of Performance Measures (PM)	
Use of Performance Information (PI)	

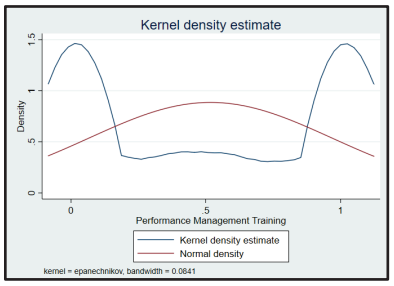
Independent Variables

Oversight (O)	
Goal Clarity (G)	
Empowered and Accountable Culture (C)	

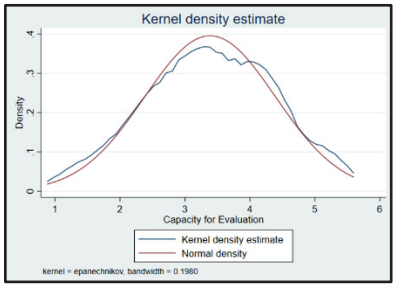
Credible Leadership Commitment (L)



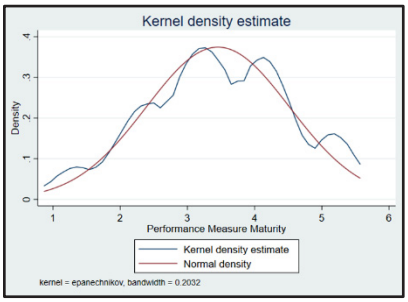
Performance Management Training (T)



Capacity for Performance Evaluation (X)

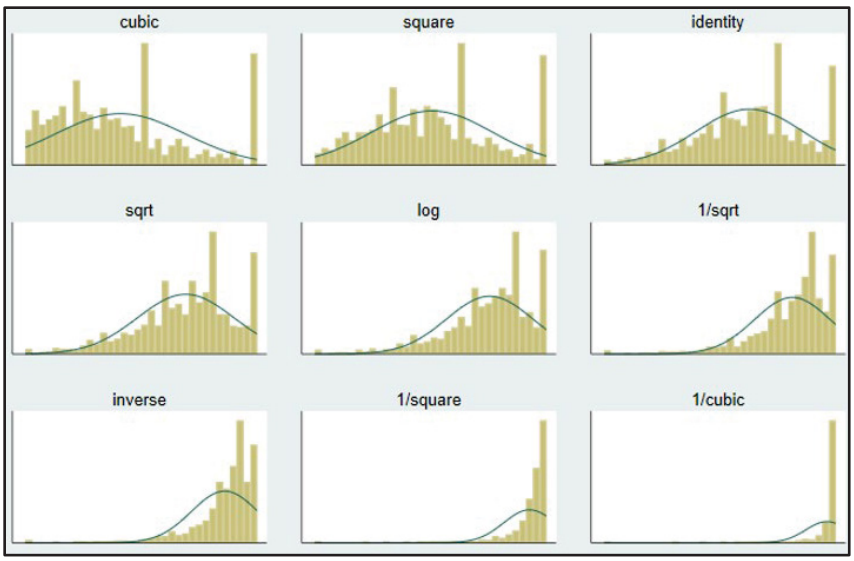


Measurement System Maturity (M)



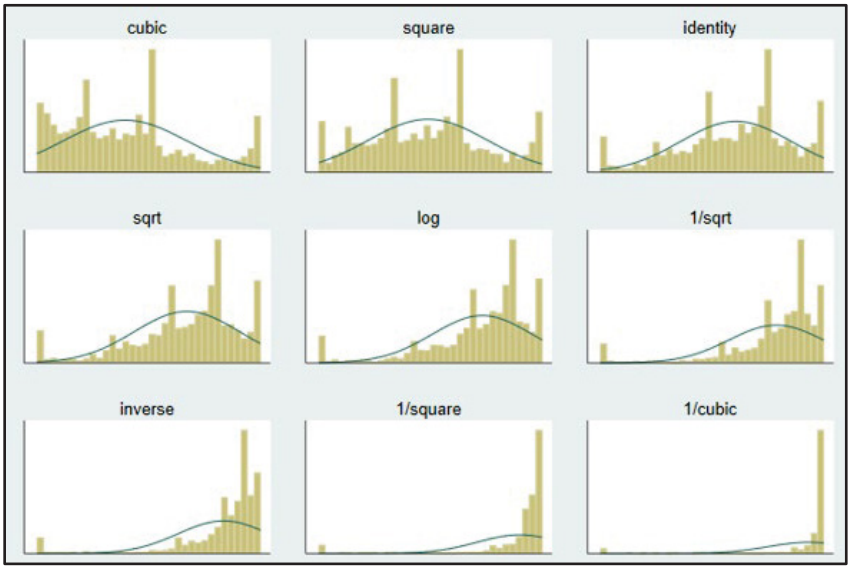
In the case of PM, Tukey's ladder of powers test indicated no transformation was required, as the identity function had the lowest chi-square value (45.65). Figure 4.3 provides a graphical impression of the effects on the normal distribution of different transformation options for PM. The visualization confirms the outcome of the ladder test.

Figure 4.3 Transformation Options for Adoption of Performance Measures (PM)



For PI, Tukey's ladder test also indicated no transformations were required (chi-squared 59.69), and the visualization in figure 4.4 confirms it as well:

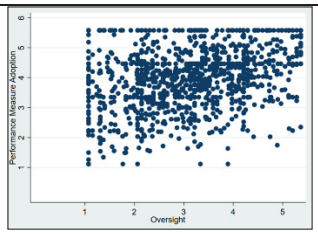
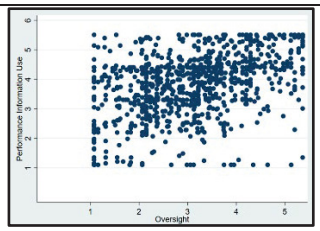
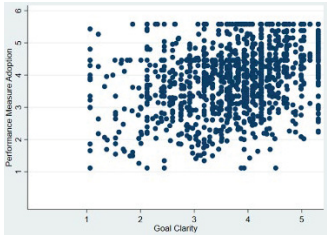
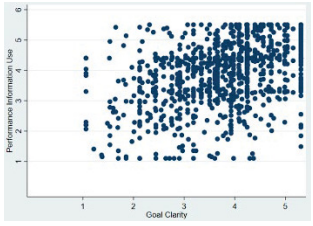
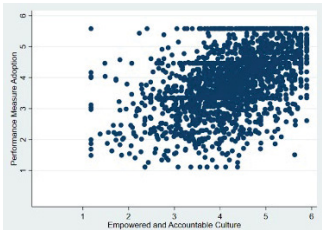
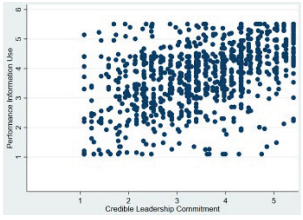
Figure 4.4 Transformation Options for Use of Performance Information (PI)



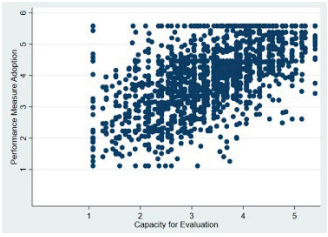
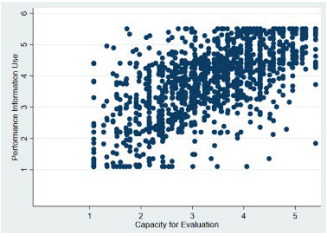
I also performed ladder tests for all independent variables. Oversight (O) was made more normal by a square root transformation. Empowered and accountable culture (C) was transformed using a squared transformation. Credible Leadership Commitment (L) was transformed using square root. Goal clarity (G), employee training (T), capacity for evaluation (X), and maturity of

performance measures (M) were not transformed. Using the non-transformed scale to support interpretability, scatterplots of the data for the independent variables with the two dependent variables PM and PI are at figure 4.5. They provide a visualization of the undeniably linear relationship between all predictors and PM and PI.[60]

Figure 4.5. Scatterplots of Predictors and Dependent Variables

	Reform Institutionalization Indicator/Outcome	
Predictor	**Adoption of Performance Measures**	**Use of Performance Information**
Oversight		
Goal Clarity		
Empowered and Accountable Culture		
Credible Leadership Commitment		

[60] The scatterplot for the factor for training appears more dispersed due to the underlying dichotomous character of the GAO survey observations.

Performance Management Training		
Capacity for Performance Evaluation		
Measurement System Maturity		

Notes:
Variable data is non-transformed and non-standardized.

A cursory examination of the scatterplots suggests the linear correlation is stronger for some predictors that for others. That conforms with the assumption of this study, as supported by the historical evidence in chapter 3, that such variance related to the relationship of system action and organizational action situation conditions with reform institutionalization exists government-wide.

As noted, O, C, and L were transformed to make their distributions more normal, a basic requirement for valid ordinary least squares (OLS) linear regressions. In order to compare and improve the interpretability of results of regression coefficients after fitting models, I rescaled all variables into Z scores to generate standardized coefficients based on the number of standard deviations, instead of ordinal survey response levels, as recommended by the UCLA Institute for

Digital Research and Education.[61] Descriptive data for the transformed and standardized latent variables is at table 4.4. The letter "z" in the variable name was added as a marker of standardization. For all constructs combined, the mean is zero and the standard deviation is 1.

Table 4.4. Descriptive Data for Standardized Transformed Latent Variables

Construct	Variable	N	Minimum	Maximum
Adoption of Performance Measures	zPM	2151	-2.74	1.63
Use of Performance Information	zPI	1643	-2.45	1.59
External Oversight	zO	1275	-2.06	1.75
Goal Clarity	zG	1607	-2.8	1.5
Empowered and Accountable Culture	zC	2386	-2.3	2.07
Credible Leadership Commitment	zL	1464	-2.6	1.49
Performance Management Training	zT	2603	-1.15	1.16
Capacity for Evaluation	zX	2021	-2.29	1.99
Measurement System Maturity	zM	2338	-2.23	1.79

Notes:
Values are z scores. These measure the number of standard deviations from the mean.

Table 4.4 shows clear variance in both the number of observations and range across the latent standardized variables, though the remaining sub-samples are perfectly adequate for drawing inferences using the analytic models. The variance in number of observations occurred as a result of the effect of accumulation of missing data from each of the GAO survey questions that were integrated into the latent variables during factor analysis. Specifically, if any associated survey question was missing data for an observation, either at random or planned, then the latent variable was also missing data for the same observation. The result of this accumulation effect was a substantial rate of missing data in the latent variables, as shown in table 4.5 (N equals 2,751).

Table 4.5. Rate of Missing Data in Latent Variables: A First Look

Construct	Variable	Missing	Missing Rate
Adoption of Performance Measures	zPM	600	21.8%

[61] https://stats.idre.ucla.edu/stata/faq/how-do-i-standardize-variables-in-stata/

Use of Performance Information	zPI	1108	40.3%
External Oversight	zO	1476	53.7%
Goal Clarity	zG	1144	41.6%
Empowered and Accountable Culture	zC	365	13.3%
Credible Leadership Commitment	zL	1287	46.8%
Performance Management Training	zT	148	5.4%
Capacity for Evaluation	zX	730	26.5%
Measurement System Maturity	zM	413	15.0%

A major uncertainty is what effect such a high rate of missing observations in some of the variables would have on modeling results. With so much missing data across all latent construct variables, the cumulative effect of so much missing data in single variables would lead to a complete case analysis of only 543 total observations after listwise deletion of observations with missing data, which could cause potential scale risks or selection effects. Therefore, I decided to fit the models two ways: first, using the dataset as is (i.e., complete case analysis), and then using multiple imputation (MI) to see if imputing the randomly-missing data would make a difference in the modeling results. In the second case, non-imputed observations with *planned* missing data would be skipped by Stata using listwise deletion. Otherwise, no skips would occur and all observations with observed data or imputed randomly-missing data would be included in the regression, thereby increasing the number of observations to 1,598 in model 1 and 1,794 in model 2.

With expert assistance from two statisticians in the Virginia Tech Statistical Applications and Innovations Group (SAIG), I prepared the dataset for MI by first observing that much of the missing data were not missing at random. As noted, either survey instructions directed respondents to skip some survey questions (i.e., planned skips), or I chose to exclude them from the factor analysis due to respondents providing a "zero" response ("Do not know/Not applicable"). Both situations effectively reduced the actual sample size to only those participating federal managers to whom the questions were salient or relevant based on GAO

instructions by excluding observations that met these criteria from the sample (including them in the sample inaccurately inflated the percent missing in table 4.5). To avoid imputing observations with "hard" missing data (i.e., for questions that were skipped due to survey guidance or given a zero by respondents), I coded the observations in the latent variables to reflect this distinction in the kinds of data missingness associated with their component survey observations. If all associated survey questions for the variable were coded for a category of planned skip or for a "Do not know/No applicable" response, then the MI variable was similarly coded for those observations. If an observation had randomly missing data for any of the questions linked to the latent variable, then I did not recode the data in the latent variable observation as planned missing and left it as missing at random for the MI process. This approach was conservative; it ensured that the MI analysis would not exclude any observations with either valid data or data missing at random from the GAO survey from the effective sample. Observations for latent variables were thus recoded per table 4.6 to place a marker that Stata would recognize into the dataset to exclude the observation from the sample to be imputed, thus not counting it as missing data. To differentiate the variable names for the MI-based regressions from the non-imputed base models, I added an "x" at the end of the variable name.

Table 4.6. Categorization of Planned Missing Data in Latent Model Variables

Variable	Recode	Reason	Condition
zPMx	.a	GAO-directed skip	When all series 5 questions are coded as directed skip (.a)
zGx	.a	GAO-directed skip	When all association series 17 questions are coded as planned skip
zGx	.b	GAO-directed skip	When all series 17 questions are coded as .b due to NRC exclusion
zMx	.c	"Not applicable/Don't know" responses	When all series 8 questions are coded as .c due to "Don't know/Not applicable" response

zPMx, zPIx, zOx, zGx, zCx, zLx, zXx	.c	"Not applicable/don't know" responses	When all associated questions are coded as .c

After applying the above rules to code these latent variables for MI analysis, I calculated the number of observations that had data missing at random[62] which would be imputed, and the number of observations which would be skipped in the MI process due to "planned" missing data (table 4.7).

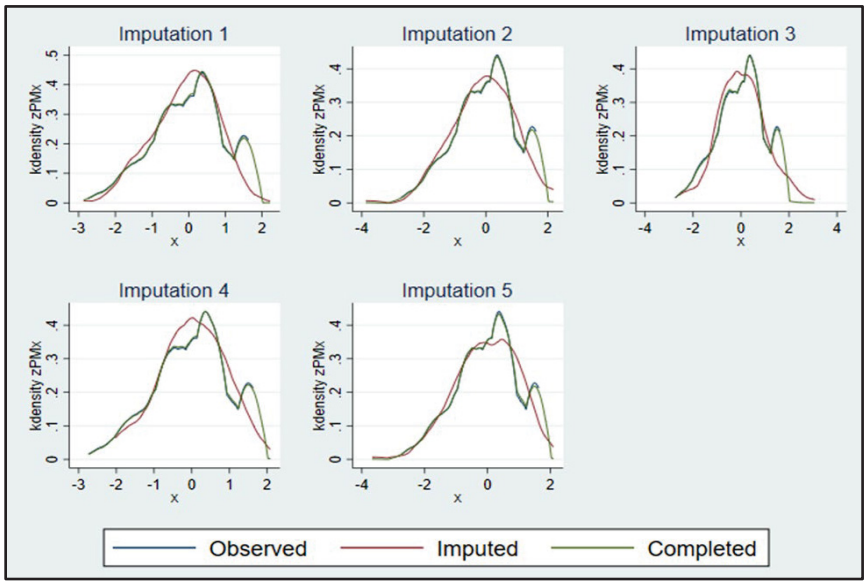
Table 4.7. Rate of Missing Data in Latent Variables and Imputable Observations

		Observations For Multiple Imputation				
Construct	**Variable**	**N**	**With Data**	**Imputable**	**To Skip**	**Missing Rate**
Adoption of Performance Measures	zPMx	2,414	2,151	263	3	10.9%
Use of Performance Information	zPIx	2,677	1,643	1,034	1	38.6%
External Oversight	zOx	2,222	1,275	947	5	42.6%
Goal Clarity	zGx	2,205	1,607	598	546	27.1%
Empowered and Accountable Culture	zCx	2,709	2,353	356	42	13.1%
Credible Leadership Commitment	zLx	2,684	1,464	1,220	67	45.5%
Performance Management Training	zTx	2,751	2,603	148	0	5.4%
Capacity for Evaluation	zXx	2,706	2,021	685	45	25.3%
Measurement System Maturity	zMx	2,635	2,338	297	116	11.3%

[62] To verify whether or not missing data for model variables were missing at random, I employed a technique proposed by Schwartz-Soicher (2017) that tests for a statistically significant relationship between any other model variables and a generated variable that includes the missing data pattern of the test variable using logistic regression models. Is a statistically significant relationship between even one model variable and the missing data pattern exists, then the missing data are missing at random (MAR). My research found this to be true for all model variables except for credible leadership and performance management training, which were missing completely at random (MCAR).

I also needed to decide on parameters for the multiple imputation operation. Using a procedure recommended by von Hippel (2020), I calculated that my models would require 136 imputations to achieve a stationary convergence of imputation distributions that produced reliable and valid estimates to replace so much missing data in the model latent variables. Each imputation generates a unique alternative dataset, and the MI procedures converge the distributions for all selected variables using a chosen method. An example of what this looks like for the first five imputations is at figure 4.6.

Figure 4.6. Example of Five Imputation Distributions for Performance Measure Adoption



For the MI operation in this study, I chose to use multivariate normal regression to fill in the missing values because, as Stata notes, "it accommodates arbitrary missing value patterns."[63] As I showed, these exist in both the model variable and the GAO datasets. To fill in missing data, Stata indicates their tool for this form of MI employs the Markov Chain Monte Carlo method.

The next section will describe the units of analysis and the hierarchical linear (mixed effects) models employed to estimate correlations of O, G, C, L, T, X and M with both PM and

[63] https://www.stata.com/manuals/mimiimputemvn.pdf

PI in their standardized (z score) forms, while controlling for the influence of the other predictors, and describe the generalized ordered logistic regression models used to estimate the correlation of the standardized latent predictors with the adoption of *specific kinds of performance measures* (e.g. outputs, outcomes, etc.) and *specific uses of performance information* by federal managers (e.g. setting goals, taking corrective actions). In addition, results of testing for violations of OLS regression assumptions will be presented. The sections following that will present the modeling results and findings using both imputed and non-imputed data. Assumptions and limitations of the quantitative modeling approach are discussed in chapter 6.

Units of Observation and Analysis and Measurement Models

The first research question was "how do external oversight, political influences, and resource dependency, as well as organizational conditions affect agency adoption of performance measures and use of performance information for management decisions?" To contribute to the debate on antecedents of performance measures adoption and performance information use from a comparative perspective, the study focuses on this relationship at the government-wide level unit of analysis. To contribute useful knowledge for practice and potential improvements in future reform governance, the study also examines these relationships at the agency level. In this study, the influence of political and resource dependency, treated as a combined concept, is modeled using comparative qualitative methods described in the qualitative research design section later in this chapter because the GAO dataset did not contain suitable questions that could serve as proxies for operationalization. In both the quantitative and qualitative portions of the study, the unit of observation is the individual federal manager and his or her perceptions.

To provide an answer to the first research question concerning the effects of external oversight and the organizational conditions including agency identity (as noted, its combined outcome and output performance orientations), the quantitative analysis employed hierarchical linear (or mixed effects) modeling to estimate the effects of these latent variables on the two indicators of GPRAMA institutionalization, the adoption of performance measures (model 1) and the use of performance information (model 2). In addition, it applied a generalized ordered logistic regression model (partial proportional odds) (model 3) to regress the adoption of specific categories of performance measures and specific uses of performance information by federal managers, expressed as ordered categorical data, with the same latent predictor variables used in the mixed effects models. To ensure model results are not biased by the current effects of past investments in staff, information technology, or performance measurement procedures, and to help control for the likely presence of endogeneity, which Baldwin, Chen, and Cole (2019) claim is common in studies of organizations where institutional influences pervade every process and actor (911), the model controls for the capacity for evaluation and measurement system maturity, both of which are autoregressive systems and proxies for past management decisions (e.g. O'Toole and Meier 2011, 29, 61).

Use of a mixed effects model that also incorporates random effects, somewhat unconventional in the study of public administration but common in other fields especially education research and organization studies, enables cross-agency/between-group comparisons not obtainable using a fixed-effects model. Use of a mixed effects model controls for the high probability that managers clustered in the same agency are likely to share substantial perspectives and be subject to the same institutional and organizational conditions and pressures, and so develop similar biases that affect their responses to survey questions. A mixed effects

model permits use of observational-level data from individual managers, while helping the research to deal with the probability that in such a survey-based study, a fundamental assumption of linear regression, i.e. that all observations are independent of each other, will be violated (Hicklin 2010, 257). It also provides the benefit of isolating and measuring the source of much unobserved variance generated by agency-unique unobserved characteristics, permitting the construct of "agency identity" to be measured in a way that allows between-agency comparisons.

As stated in chapter 1, the research purpose is to generate findings of practical value for public managers interested in changing their agency's governance environment or organizational strategies, and to contribute empirical findings to the growing literature on purposeful uses of performance management in government agencies. It is not to analyze conditions at the micro-level in order to address tactical management problems or recommend specific changes to business processes to improve performance in individual agencies. Therefore, the quantitative analysis has a modest objective to test the statistical and substantive significance of all theorized relationships to validate claims in the literature, not to report precise point estimates or slopes (on a case-by-case basis, such micro-analysis could prove a fruitful and valuable contribution for future research). Predictor coefficients of the fixed effects and random effects parameters for models 1, 2, 4, and 5 and odds ratios for model 3 predictors are designed therefore to reveal the relative level of influence of the predictors on the dependent variables government-wide and at the agency level.

I used two sample-weighted hierarchical linear (mixed effects) regression models that estimate fixed effects for external oversight and the organizational condition variables from the

inferential model.[64] Under this approach, agency identity, the final organizational condition in the inferential model, is estimated as a second-level random effect both as a control for the fixed effects variables and to measure its separate influence on GPRAMA institutionalization indicators to permit agency-level comparisons across the sample. Chapter 5 gives a fuller account of how this was done as part of an expanded discussion of the quantitative research design used to answer research question 2.

To analyze the relationships of the system action and organization action predictors with the latent dependent variables of PM and PI, the study uses Stata version 16 software to fit two models. Model number 1 uses the Stata MIXED command to fit a bi-level mixed effects weighted and agency-clustered model to estimate the correlations and significance between all predictor variables and PM. Model number 2 replaces the dependent variable in model 1 with PI, but is otherwise identically fitted. Figure 4.7 presents models 1 and 2.

Figure 4.7. Hierarchical Linear Models for GPRAMA Institutionalization

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{PMij}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 O_{ij}} + \boldsymbol{\beta_2 G_{ij}} + \boldsymbol{\beta_3 C_{ij}} + \boldsymbol{\beta_4 L_{ij}} + \boldsymbol{\beta_5 T_{ij}} + \boldsymbol{\beta_6 X_{ij}} + \boldsymbol{\beta_7 M_{ij}} + \boldsymbol{z_{ij} u_j} + \boldsymbol{\varepsilon_{ij}}$$

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{PIij}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 O_{ij}} + \boldsymbol{\beta_2 G_{ij}} + \boldsymbol{\beta_3 C_{ij}} + \boldsymbol{\beta_4 L_{ij}} + \boldsymbol{\beta_5 T_{ij}} + \boldsymbol{\beta_6 X_{ij}} + \boldsymbol{\beta_7 M_{ij}} + \boldsymbol{z_{ij} u_j} + \boldsymbol{\varepsilon_{ij}}$$

$\widehat{\boldsymbol{Y}}_{\boldsymbol{PMij}}$ is the estimated agency adoption of performance measurements across observations i and agency clusters j
$\widehat{\boldsymbol{Y}}_{\boldsymbol{PIij}}$ is the estimated agency use of performance information for management decisions across observations i and agency clusters j
$\boldsymbol{\beta}$ are fixed effects regression coefficients
$\boldsymbol{x_{ij}}$ are the fixed effect predictors (O=Oversight, G=Goal Clarity, C=Culture, L=Leadership, T=Training, X=Capacity, and M=Maturity)
$\boldsymbol{z_{ij}}$ are covariates of the random effects with the dependent variable
$\boldsymbol{u_j}$ are the random effects of agency identity
$\boldsymbol{\varepsilon_{ij}}$ are random errors

[64] Because all of the quantitative models use sample weights that account for differences in the size of agency manager cohorts to enable generalization across the federal government, all point estimates are reported using robust standard errors.

To regress agency adoption of particular kinds of measures and specific uses of performance information with the system action and organization action situation predictors, model 3 is fitted using a single-level partial proportional odds model[65] implemented with a customized Stata module, GOLOGIT2, developed by Richard Williams at the University of Notre Dame (R. Williams 2006). This is a change to the originally-intended method of using a hierarchical mixed effects model using a mixed effects ordered logit model. The change was necessary because careful analysis of the underlying GAO data using the Brandt test revealed that the ordinal survey data associated with several model variables violated the parallel lines/proportional odds test.[66] Without being able to assume proportional odds, using an ordered logistic regression model--mixed or not--would have generated invalid results. An ordered logistic regression model is simply a special case of a generalized ordered logistic model, one that may be employed when the underlying data do not violate the Brandt test.

The partial proportional odds model measured the odds that a level change in a predictor variable would increase the level of manager response to the dependent variable survey question of interest from level 3 ("agree to a moderate extent") to level 4 ("agree to a great extent"). Williams (2006) argues a generalized ordered logit model is ". . . more parsimonious and interpretable than those fitted by a nonordinal method, such as multinomial logistic regression" (58). Fullerton and Xu (2012) further note that in the family of generalized ordered logit models, a partial proportional odds model keeps the proportional odds assumption for variables where it

[65] A partial proportional odds model relaxes the proportional odds assumption to allow comparison of effect estimates at specific levels (R. Williams 2006).
[66] I ran the Brant test using on the logistic regression model using two dependent variables without weights or a second level cluster, both required exclusions for the test. Culture, training, and capacity for evaluation were statistically significant, thereby leading me to reject the null hypothesis of parallel lines/proportional odds between ordered levels. This precluded the use of either ordered logit models or mixed effects ordered logit models.

holds true empirically, and relaxes it for those in which it does not, and so is more efficient than generalized ordered logit models which relax the assumption for all variables (184).

Using this methodology for model 3, the predictor's coefficient, expressed as an odds ratio, expresses the odds that a one-level survey scale increase in the predictor variable will increase the institutionalization of the GPRAMA practice measured by the dependent variable from level 3 to level 4, i.e., to the point where managers perceive with minimal uncertainty that the practice has become institutionalized in their agency. The transition from level 3 to 4 as the threshold for evidence of institutionalization of GPRAMA practices was selected on empirical and intuitive grounds. The differential ambiguity of organizational conditions between "modest agreement" and "great agreement" seems to unambiguously signal manager perceptions that the practice has been routinized in the agency, so a response of "very great" is taken to sufficiently affirm a federal manager's perception that the condition measured by the question has achieved a significant degree of institutional sedimentation in their agency.[67] Figure 4.8 presents model 3 derived from guidance by Fullerton and Xu (2012, 184).

Figure 4.8. Partial Proportional Odds Model for Estimating Specific Practice Institutionalization

$$\ln\left(\frac{Pr\{y \leq m_{ij}|x\}}{Pr\{y > m_{ij}|x\}}\right) = \tau_{m_{ij}} - X_{ij}\boldsymbol{\beta} - \boldsymbol{\omega_{ij}}\eta_m \; (1 \leq m < M)$$

Where:
m is the ordered response category (i.e. PM or PI)
X and **ω** are vectors of the independent variables O, G, C, L, T, X, and M
τ is a cut point (here, 4)
β is a vector of logit coefficients that are fixed across cut point equations
Ƞ is a vector of logit coefficients that vary across cut point equations
M is the highest response category
ij implies across observations i and agency clusters j

[67] A possible empirical source of validation for this threshold is how GAO uses it in its periodic reports about GPRAMA; the threshold between 3 and 4 regularly appears as the basis for analytic differentiations between federal agencies in terms of their implementation of GPRAMA requirements.

Model Testing and Validation

I tested models 1 and 2 for the existence and extent of OLS assumption violations (model 3 is not OLS). I first tested for the degree of multi-collinearity. Excess collinearity between model variables can pose a threat to the validity of research by affecting the sign or inflating estimates of the variance of parameters, and other issues (O'Brien 2007, 673) . Not unexpected for a survey instrument measuring human perceptions of similar phenomena, pairwise correlations in figure 4.9 showed a degree of collinearity between all variables (all correlations were significant at the .01 level).

Figure 4.9. Pairwise Correlations of Model 1 and Model 2 Variables

	PM	**PI**	**O**	**G**	**C**	**L**	**T**	**X**	**M**
PM	1.000								
PI	.612*	1.000							
O	.341*	.402*	1.000						
G	.405*	.444*	.398*	1.000					
C	.539*	.616*	.472*	.546*	1.000				
L	.437*	.491*	.734*	.569*	.665*	1.000			
T	.314*	.362*	.278*	.261*	.355*	.329*	1.000		
X	.611*	.636*	.502*	.537*	.721*	.655*	.432*	1.000	
M	.341*	.285*	.114*	.264*	.296*	.233*	.153*	.432*	1.000

Notes:
* p<.01

I tested both models for the variable inflation factor (VIF) after performing an OLS regression with sample weights but without MI. O'Brien (2007) states "the VIF indicates how much the estimated variance of the *i*th regression coefficient is increased above what it would be if *[the pairwise correlation] eq*ualed zero" (674). Results at table 4.8 are within conventional norms and indicate the collinear data are unlikely to cause significant inflation of predictor coefficients.

Table 4.8. Variable Inflation Factor Scores for Model 1 and Model 2 Independent Variables

Predictor Variable	VIF Model 1	VIF Model 2
O	2.59	2.55
G	1.69	1.7
C	2.6	3.08
L	3.36	3.11
T	1.47	1.5
X	2.73	3.25
M	1.36	1.4
Mean VIF	2.26	2.37

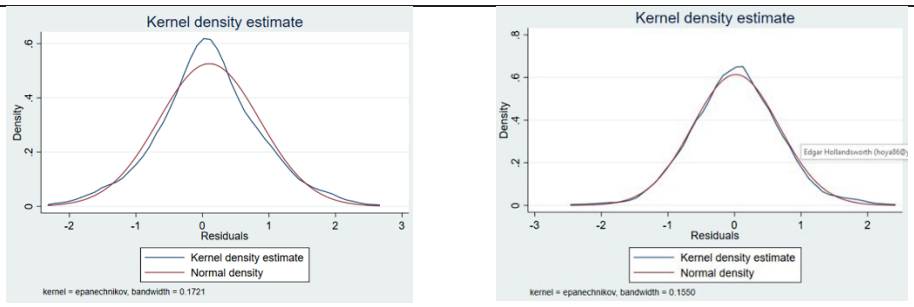
According to O'Brien (2007), the statistical literature commonly cites a VIF of 10 or higher as cause for concern, although some scholars suggest 4 as a rule of thumb (674). As all VIF values in these models are below even the more demanding threshold, I assume multi-collinearity exists but does not pose an unacceptable risk for the validity of the regression coefficients.

I also tested for the normality of model residuals. OLS regression requires residuals to be distributed independently and normally to ensure that p values in the t-test are valid. For models 1 and 2, the distributions appear fairly normal but with some kurtosis evident, especially for model 1. Figure 4.10. presents the distributions and normal probability plots of observed versus expected normal values.

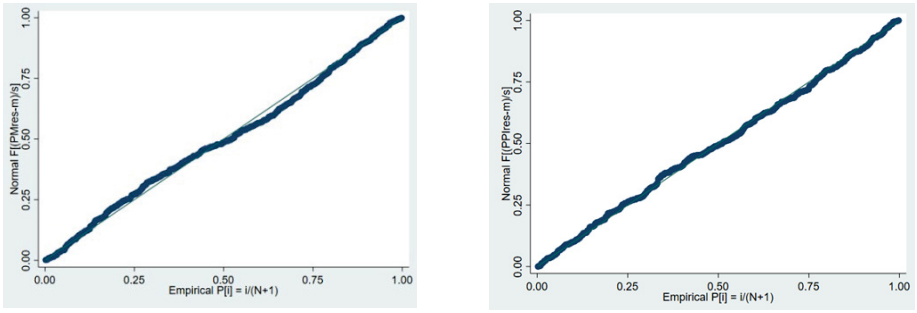
Figure 4.10. Evidence of Normal Distribution of Model 1 and 2 Residuals

Test of Normality	**Model 1**	**Model 2**

Normal Kernal Density Chart



Normal Probability Chart



Because figure 4.10 showed some degree of deviation from the normal distribution for model 1, I ran the Shapiro-Wilk test. The p-value was .0385, too high to reject the null hypothesis that the distribution is normal at the .01 level, but low enough to reject it at the .05 level. So, I conclude the distribution of residuals for model 1 is near-normal.

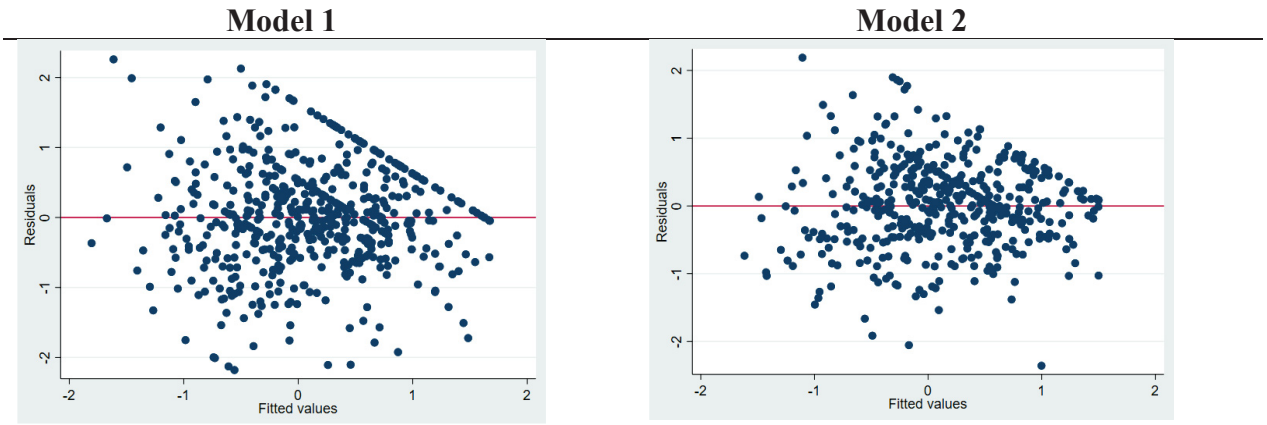
In this case, I investigated whether or not the existence of outliers in the dataset might explain the minor deviation from perfect normalcy. I downloaded and installed the IQR module for Stata developed by Hamilton (Hamilton 1992) to identity the number and severity of outliers that might be causing the distribution of residuals to be less than perfectly normal. For the set of residuals (N=543), the mean was .102, the standard deviation was .758, the number of mild outliers at the low end (more than -1.717 standard deviations from the mean) was 5 (.92% of the sample) and the number of mild outliers at the high end (more than 1.918 standard deviations) was 8 (1.47% of the sample). There were no severe outliers (<-3.08 or >3.28 standard deviations). For model 2, Shapiro-Wilk produced a p value of .248, not sufficient to reject the null hypothesis of normal distribution. I conclude the residuals are sufficiently normal, given the

very small percentage of outliers and the use of underlying data which was ordinal originally with a fixed range for the latent variables. The residual distribution seems to be as normal as a model with constructed factor variables based on Likert scale survey responses is likely to be.

Models to test for homoskedasticity are sensitive to even small deviations from normality in the distribution of residuals. It is therefore not too surprising to discover heteroskedasticity in the model data, for which I likewise tested. The Breusch-Pagan test demonstrated moderate levels of heteroskedasticity for models 1 (Chi^2 =21.64, p>.001) and 2 (Chi^2=27.64, p>.001). An information matrix test also confirmed the presence of heteroskedasticity (Model 1: Chi^2=78.43, p>.001; Model 2: Chi^2=87.88, p>.001) as well as kurtosis (Model 1: Chi^2=7.50, p>.01; Model 2: Chi^2=6.75, p>.01).

Plots of the residuals at figure 4.11 show the residuals in both models exhibit a pattern of declining variance across the range of fitted values. The clear linear decline in highest residual values at every level as well as several lines parallel to that strongly suggest the form of the data which was originally Likert scale prior to polychoric correlation and factor analysis is responsible for the pattern's appearance. Thus, some degree of heteroskedasticity is an outcome of using a set of latent values for ordinal survey data, and was obviously unavoidable.

Figure 4.11. Plots of Residuals for Models 1 and 2



Nevertheless, the errors are not independent and identically distributed. Cameron and Trivedi (2010) note that ". . .in much applied work. . .heteroskedasticity is the norm. . ." and recommend the use of robust standard errors as a means to generate heteroskedasticity-robust estimates (84). The default setting for mixed effects models in Stata uses robust standard errors. This study will simply accept a degree of instrument-imposed heterogeneity and try to mitigate its effects on the results using robust standard errors. Not needing precise point estimates to answer the research question helps the study to make such an accommodation.

Next, I tested for linearity and specification sufficiency. Augmented component-plus-residual plots demonstrated linear relationships between all predictors and dependent variables in the same way the bi-variate scatterplots did. To test model specifications, link tests showed the prediction squared offered no additional explanatory power over either of the base models. Omitted variable (regression specification error) test results (for model 1, $F=3.24$, $p=.022$; for model 2, $F=2.11$, $p=.0989$) showed model specification errors are significant at the .01 level for neither model, but are significant at the .05 level for model 1. This suggests the models are sufficiently specified, but additional unknown variables or perhaps small changes in the model predictors could explain more of the variation in the dependent variable for model 1. The study seeks only to put the major claims in the public management literature to the test, not to integrate every conceivable factor that might affect PM or PI. The need for parsimony must be balanced against the need for complete specification. The literature review showed a wide range of alternative possible predictors that simply could not be operationalized into the variables of the current model in a meaningful or parsimonious way. In addition, the risks of miss-specification should be at least partially mitigated by use of agency clustering, sample weighting in all models, robust standard errors, and inclusion of random effects in the hierarchical linear models which

absorb unobserved agency-specific variance into agency identity that would otherwise bias fixed effects point estimates.

Quantitative Findings

The results for models 1, 2, and 3 (excluding agency identity, discussed in chapter 5) are shown in table 4.9. Only statistically significant results at the .01, 05, and .10 levels, as noted, are displayed. Base models 1 and 2 used MI to replace missing data in the latent variables generated during the variable construction process using factor analysis, as described previously. As an excursion, and as described in the research design, I also fitted models 1 and 2 without using MI as a robustness check. Table 4.10 shows the results for this variant. All coefficients are estimated using robust standard errors.

Based on the use of standardized data for the latent variables, my interpretation of the coefficients for models 1 and 2 is that they are the percentage of one standard deviation in the variance of the dependent variable (in bold) that is associated with a 1 standard deviation increase in the predictor variable. Coefficients for model 3 dependent variables (in italics) are expressed as odds ratios, so these are interpreted as the odds that a one standard deviation increase in the predictor variable will raise an agency's average survey score from level 3 (proxy for manager-perceived partial institutionalization) to level 4 (proxy for manager-perceived sedimented institutionalization). Odds ratios greater than one suggest a positive association; those less than one, a negative association.

Table 4.9. Results of Models 1 and 2 with Imputation, and of Model 3 Without Imputation

		Predictor Variables						
Dependent Variables	**N**	**Oversight**	**Goal Clarity**	**Culture**	**Leadership**	**Training**	**Capacity**	**Maturity**
Performance Measure Adoption &	1595			.160 (.051)**		.127 (.042)**	.327 (.087)**	.188 (.069)**
Output Measures	565			1.613 (.186)**			1.729 (.310)**	1.322 (.109)*
Efficiency Measures	570			1.870 (.259)**	.717 (.088)**		2.447 (.414)**	1.351 (.130)**
Customer Service Measures	565			1.593 (.227)**			1.696 (.260)**	1.489 (.236)*
Quality Measures	567			1.897 (.285)**			1.776 (.267)**	1.429 (.109)**
Outcome Measures	571			2.123 (.267)**			1.925 (.301)**	1.537 (.147)**
Performance Information Use &	1737			.211 (.048)**			.275 (.102)**	
Program Strategy	601			2.257 (.373)**	1.744 (.341)**		1.838 (.323)**	
Program Priorities	603			2.070 (.242)**	1.517 (.231)**		1.546 (.210)**	.
Resource Allocation	596			2.190 (.262)**	1.418 (.175)**		1.522 (.260)*	
Problem Identification	601			2.173 (.350)**	1.646 (.313)**		1.870 (.331)**	
Corrective Actions	604	.712 (.119)*		2.449 (.509)**	1.474 (.261)*		1.759 (.289)**	1.188 (.092)*
Work Processes	598		1.181 (.112)+	1.794 (.298)**			2.039 (.395)**	
Performance Goals	598		1.295 (.130)**	1.641 (.220)**	1.546 (.258)**		1.958 (.346)**	
Contract Management	483	1.545 (.175)**		1.256 (.166)+		.873 (.065)+	2.263 (.344)**	
Program Consolidation	586	1.401 (.176)**		1.554 (.207)**			2.361 (.381)**	1.286 (.182)+

Notes:
** p<.01, * p<.05, + p<.10. Robust standard errors in parentheses.
& Sample size increased via multiple imputation
Coefficients for survey question variables expressed as odds ratios to aid interpretability.

Only statistically-significant coefficients displayed.

Table 4.10. Results of Models 1 and 2 without Multiple Imputation

		Predictor Variables						
Dependent Variables	**N**	**Oversight**	**Goal Clarity**	**Culture**	**Leadership**	**Training**	**Capacity**	**Maturity**
Performance Measure Adoption	543		.054 (.030)+	.213 (.086)*		.090 (.050)+	.409 (.094)**	.213 (.086)*
Performance Information Use	455						.407 (.065)**	

Notes:
** p<.01, * p<.05, + p<.10. Robust standard errors in parentheses.
Coefficients for survey question variables expressed as odds ratios to aid interpretability.
Only statistically-significant coefficients displayed.
& Sample size increased via multiple imputation

The differences between the results of the two variants for models 1 and 2 are few, suggesting the models were fairly robust to the use of MI and the expanded size of the sample. The differences, however, should be noted as they reveal important insights on the influence of some organization action situation conditions. In model 1 variant 1 using imputed data, performance measure adoption is significantly influenced by an empowered and accountable culture (C), availability of performance management training (T), capacity for evaluation (X), and maturity of the measurement system in the organization (M). Model 1 variant 2 (non-imputed) produced similar results, but added goal clarity (G) as a significant predictor, but only at the .10 level and with an insubstantial coefficient. In addition, the significance level for T fell to .10 and its coefficient fell from .127 to .090, making it less substantively important. Therefore, the use of MI generated somewhat different results for model 1 than using complete case analysis and extensive listwise deletion. I therefore find that a culture of empowerment and accountability, the capacity for evaluation, and the maturity of the performance measurement system are all positively associated with the adoption of performance measures. External

oversight, goal clarity, and leadership have no statistically significant association with performance measure adoption in the U.S. Government, when controlling for the influence of other organizational conditions. While training in performance management is statistically significant, the small regression coefficient and the lack of a single significant association with the adoption of any performance measure leads me to conclude it is not analytically substantial.

In model 2 variant 1 (imputed), C and X are statistically significant, at the .01 level. In variant 2, only X is significant. As X and M are considered autoregressive control variables in this study, this means none of the other five system or organization action conditions are significant if the data is not imputed. Thus, it appears use of MI adds value to the study by helping to reveal an association between C and PI that would have been overlooked without its use. In addition, the results show both C and X have a statistically significant and uniformly positive correlation with every form of PI use in the study. Credible leadership commitment was not significant as an antecedent for PI use generally, but demonstrated a significant and positive correlation with six of nine distinct uses of PI. Based on these results, I find that an agency culture of empowerment and accountability and the capacity for performance evaluation are positively associated with PI use. Credible leadership commitment may play an important role for specific uses of PI in federal agencies.

Because X and M are autoregressive control variables, their covariational influence was predicted to be strong based on theoretical propositions and empirical evidence in the literature review. Based on the degree of collinearity between predictors, which as noted in the previous section does exist but should not pose a serious threat to validity, I suspected their influence on PM and PI might be so strong as to hide or wash out meaningful covariance of the other predictors of interest. I therefore ran a second excursion, using MI, that excluded X and M from

models 1 and 2. The results are shown in table 4.11. Removing the two controls exposed a statistically significant relationship for goal clarity with both indicators of GPRAMA institutionalization, more so with the adoption of performance measures.

Table 4.11. Results of Models 1 and 2 with Imputation but Excluding Capacity and Maturity

		Predictor Variables				
Dependent Variables	**N**	**Oversight**	**Goal Clarity**	**Culture**	**Leadership**	**Training**
Performance Measure Adoption	1,598		.118 (.042)**	.353 (.045)**		.185 (.042)**
Performance Information Use	1.092	.	.067 (.040)+	.306 (.058)**		.169 (.077)*

A possible explanation for these results is that X and M incorporate clear goals as a foundation for past efforts to build organizational capacity and for developing sophisticated means for measuring performance. Once capacity and measurement methods are institutionalized, improving goal clarity makes far less difference for future performance reform outcomes. It is likely this is because goals have already been clarified. A possible implication of this proposition is that agencies which have less advanced infrastructure for evaluating performance, and have ineffective measurement systems, may advance GPRAMA institutionalization by focusing on clarifying goals for their organization and its programs, while those which have achieved the level of sedimentation in their performance management practices should focus more on building a climate of accountability and empowerment in their cultures to sustain those practices.

Model 3 results show the association of the model 1 and 2 predictors with the adoption of specific kinds of measures (e.g., output measures) and specific uses of performance information (e.g., developing program strategy). Model 3 is a generic model, in which the dependent variable is a single GAO survey question with five-level ordinal data, and multiple questions are

regressed on to the standard latent predictors from models 1 and 2. As noted in the research design, the analysis focuses on the critical threshold of the transition from level 3 to level 4 GAO survey responses as a cutoff to affirm federal manager perceptions of the sedimentation of performance management practices in their agencies. Table 4.9 presents the governmentwide results including the two controls, capacity and measurement system maturity. Oversight showed no significant correlation with adoption of any category of performance measure, and a relationship with uses of performance information for taking corrective actions (a negative relationship!), contract management and program consolidation (both positive). This comports with the record of general unresponsiveness by many federal agencies to GAO appeals for action to implement GPRAMA: most agency reports to Congress, the GAO, and other oversight authorities related to contract management and agency programs are not closely tied to GPRAMA reporting systems; rather to budgetary and acquisition processes.

In base model 3, goal clarity was related only to use of performance information for work processes and setting performance goals. One might expect that clarity of goals, especially in a strategic or performance plan, should encourage the adoption of a variety of performance measures. In many cases, it probably does. When considered governmentwide and controlling for other variables in base model 3, however, its relative importance fades. Because the excursion for models 1 and 2 showed a masking effect on goal clarity from the inclusion of the two control variables, we replicate that excursion for model 3 to see if a similar pattern emerges whereby goal clarity, or other predictors, demonstrate a stronger association with GPRAMA practices when excluding the control variables. As noted, this could be relevant for agencies with less mature institutionalization of performance management routines. Table 4.12 shows the results.

Table 4.12. Results of Model 3 Excluding Capacity and Maturity

		Predictor Variables				
Dependent Variables	**N**	**Oversight**	**Goal Clarity**	**Culture**	**Leadership**	**Training**
Output Measures	624	.811 (101)+	1.341 (.145)**	2.198 (.256)**	1.280 (.149)*	
Efficiency Measures	630			1.870 (.259)**	.717 (.088)**	
Customer Service Measures	626		1.193 (.125)+	2.275 (.230)**		
Quality Measures	630		1.234 (.124)*	2.277 (.273)**		1.262 (.146)*
Outcome Measures	571	.823 (.095)+	1.404 (.135)**	2.606 (.297)**	1.324 (.225)+	
Program Strategy	664			2.560 (.351)**	1.974 (.320)**	
Program Priorities	666		1.217 (.116)*	2.441 (.234)**	1.589 (.249)**	
Resource Allocation	657		1.246 (.136)*	2.414 (.221)**	1.455 (.157)**	
Problem Identification	665		1.196 (.115)+	2.821 (.342)**	1.557 (.345)*	
Corrective Actions	668			3.009 (.518)**	1.368 (.214)*	
Work Processes	661		1.290 (.104)**	2.307 (.266)**	1.454 (.196)**	
Performance Goals	661		1.328 (.126)**	1.975 (.252)**	1.652 (.266)**	1.227 (.115)*
Contract Management	529	1.478 (.176)**		1.651 (.185)**		
Program Consolidation	645	1.485 (.184)**		2.552 (.331)**		

Notes:
** p<.01, * p<.05, + p<.10. Robust standard errors in parentheses.
Coefficients are expressed as odds ratios to aid interpretability.
Only statistically-significant coefficients displayed.

Relaxing the constraints of the two autoregressive control variables, evaluation capacity and measurement system maturity, the model 3 excursion shows that oversight is associated with

four performance management practices, vice 3 in the base model, a minor, likely inconsequential, change. Under this model, oversight is weakly associated with agency adoption of output and outcome measures, but is no longer related significantly to taking corrective actions. The results also suggest a positive association between oversight and federal manager perceptions of the use of PI in the management of contracts and the consolidation of programs in their agencies. Through its oversight of performance management and control of the public purse, Congress and other authorities such as GAO and Inspectors General can influence management decisions in these areas quite readily; for other manager uses of performance information, less so. The overall finding of relative unimportance of external oversight was robust to the excursion.

On the other hand, the re-specified models (or, excursions) increased the number of performance management practices that are significantly associated with the latent predictor variable Goal Clarity from 2 to 9, and 7 out of the 9 were significant at the .05 level. Building on the results of the excursions for models 1 and 2 that eliminated the capacity and maturity variables which suggested an increased role for Goal Clarity, the results of the model 3 excursion simulate which specific performance management practices can be more readily institutionalized by improving Goal Clarity in an agency which has a performance management system with inadequate capability for evaluation and/or primitive performance measurement systems. These include adoption of four out of the five kinds of performance measures, and using performance information for 5 out of 9 kinds of management actions. Thus, for public managers in these kinds of federal agencies, articulating clear goals can make a real difference in their agency's institutionalization of GPRAMA and in the usefulness of performance information for daily work.

Culture has a very strong association with the adoption of all categories of performance measures, with the strongest relationship with adoption of outcome measures, in both model 3 variants. An agency with a culture of empowerment and accountability is two to two and a half times as likely to adopt outcome measures as agencies without such a culture, all else being equal. This is an important finding, because as the literature review in chapter 2 and history in chapter 3 demonstrated, an important legislative intent for GPRAMA was for agencies to adopt a cultural orientation of managing for results, or outcomes. Agencies with an empowered and accountable culture are better positioned to adopt outcome measures than those which do not have such a culture. In addition, reviewing changes in the culture coefficients for all performance management practices, the point estimate is substantially higher across the board. An interpretation of this result is that cultural change is always important for performance management reform, but even more so when an agency has made few investments in evaluation capacity or failed to develop effective performance measurement systems. Once these are institutionalized, these organizational conditions replace some of the need for emphasis on cultural improvement. Of course, this scenario would permit managers to devote more attention to matters other than cultural improvement.

In the base model 3, credible leadership commitment was a significant condition only for contract management, and only at the .10 level. Removing the two control variables, it becomes significantly associated with ten GPRAMA practices, nine of which are at the .05 level. The most likely explanation for this is that in an agency with more fully institutionalized practices for evaluating and measuring performance, credible leadership commitment is perceived by federal managers as a less salient condition for adopting new performance measures and using performance information. The implication of this is that the commitment of public leaders to

performance management practices is a critical success factor during the habitualization phase of institutionalization and, as noted in chapter 2, active leader advocacy for the reform is a necessary condition to advance institutionalization to the objectification stage. Once an agency sediments performance management practices into daily measurement routines and makes the necessary investments in evaluation capacity, federal managers may no longer perceive much impact from leadership commitment, because the agency has begun to routinely measure its performance without management engagement and active advocacy. Agency senior leaders should thus be encouraged that their credible efforts early on which succeed will allow them to redirect their limited attention to other matters once institutionalization of performance management routines becomes sedimented in their agencies.

Qualitative Research Design

Introduction

To make this study more relevant for public management practice, I sought to extend the study beyond calculating abstract estimates of correlations between latent variables in order to make inferences and original contributions using two qualitative research designs. Each is treated as a qualitative sub-study with different research and analytic methods. First, in sub-study 1, following the recommendations of Thomann and Maggetti (2020) and guided by the research plan in table 4.1, I sought to discover which configurations of necessary and sufficient conditions that were deemed to be statistically significant in the quantitative research lead to the institutionalization of GPRAMA practices in federal agencies. Understanding and interpreting the causal configurations is the basis for modest generalization about causal patterns and the further development and empirical validation or falsification of related public management and policy implementation theories.

The specific objective of sub-study 1 was to generate a limited number of set-theoretic hypotheses based on inductively-derived configurations of necessary and sufficient conditions in the system action situation and organization action situation, which were determined to be statistically and substantively significant for the institutionalization of GPRAMA, that led federal managers to perceive a sedimented (i.e. "taken for granted") level of institutionalization of GPRAMA practices in their agencies, or the absence thereof. Drawing from scholarship on configurational comparative methods, core assumptions behind the research design for this section are that configurations of necessary and/or sufficient conditions leading to agency institutionalization or non-institutionalization of GPRAMA are affected by system and organization action situation conditions in federal agencies, both by their presence and by their absence. In addition, such causal configurations are expected in some cases to be equifinal, meaning that federal agencies can achieve deep institutionalization of performance management practices through more than one possible configuration (or path) of necessary and/or sufficient conditions. As argued in chapter 1, analyzing which causal paths are associated with institutionalization may provide public managers an empirical basis to envision strategies and plans to change conditions in external and organizational environments in a way that eventually or indirectly facilitates performance management institutionalization.

The second goal of the qualitative research, pursued in sub-study 2, was to discover and document empirical evidence about the effects of system and organization action situation conditions and internal causal mechanisms (processes) within federal agencies, using a small-N cross-case synthesis research design (Yin 2014, 164-168), and analyze how the mechanisms are associated with postulated causal configurations identified by the QCA-based findings and are either present or absent across the agency cases. Ragin (2008) observed

> . . . the combination of conditions cited by the investigator should have an internal coherence – the combination should make sense as a causal recipe. *Ultimately, causation can be observed only at the case level*; a combinatorial causal argument provides explicit guidance regarding what to observe in an empirical case and very often also implies specific causal mechanisms that both link the different [causal] ingredients together and indicate the nature of their connections to the outcome. (112, italics added)

In empirical research, evidence about changes in agency processes has been found to be valuable for differentiating between mere claims of the presence of a practice and its actual use in operations (Yetano 2013, 80). In this study, the qualitative findings about configurations of causal conditions helped to verify the quantitative results as much as the quantitative results pointed the way for the investigation of specific causal mechanisms. This points to the potential for suggesting a fruitful analytic dialogue between micro- and meso-level phenomenon related to GPRAMA institutionalization facilitated by a mixed methods strategy.

To identify and characterize micro-level conditions in agencies and their effects, I collected narrative and descriptive data through interviews from six federal performance system managers (PSMs) who are responsible for developing and embedding routine practices to support agency responsiveness to GPRAMA in five federal agencies. When they develop and institutionalize agency performance management practices and routines, federal PSMs are in some ways similar to "street level" bureaucrats, though the "street level" moniker is typically applied to public-facing and service-providing public employees. In the context of GPRAMA, they qualify for this designation because they work at the leading edge of the institutionalization process and deliver performance measurement services to line managers, public leaders, and oversight authorities. As agency performance information is also shared with citizens, PSMs also provide a valuable service to individual citizens in the same way common "street level" employees do. Interviews with the agency PSMs presented an excellent opportunity to obtain data on their attitudes and biases concerning performance management and their agencies'

institutionalization of GPRAMA. The interviews captured empirical data and candid manager perspectives that are not collected elsewhere, to the best of the author's knowledge. The data gathered through the PSM interviews should make an empirical contribution to knowledge about federal PSM attitudes and biases and the relative importance of system action and organization action conditions to their decisions to institutionalize GPRAMA-related routines.

This part of the qualitative research did not attempt to systematically map or measure even a small fraction of all possible individual psychometric factors shaping public manager perceptions and decisions at the micro-level, but it did collect information regarding their intrinsic motivation and attitudes towards performance management which are part of the micro-level analysis within agency case studies. This data supplements the data from the 2021 survey for the five agencies with PSMs who chose to participate in interviews. The results of this research offer normative value for public management practice: federal PSM attitudes and biases can play an important role in the successful institutionalization of performance management reforms in agencies. While a positive attitude by a PSM is not sufficient for GPRAMA institutionalization, it pays to have an advocate running a program that he or she believes has real merit.

Data Source and Sample

Data was collected using a survey from a non-random sample of 20 federal performance managers from mid-level (GS-13) to senior executive (Senior Executive Service) who volunteered to participate in the study. Participants in interviews self-selected into the study after learning about the opportunity to participate and having their responses used anonymously. The 2021 survey data cannot be used to generalize to the population of federal employees or managers in the way the GAO survey data can. The surveys were developed on, and

administered via, the Virginia Tech Qualtrics (Qualtrics, Provo, UT) platform.[68] In addition, I conducted a series of structured interviews with a subset of six of the federal performance manager survey respondents representing five federal agencies after they had completed their surveys. The purposes were to clarify their survey responses and to ask additional questions about their attitudes towards performance management, their public service motivation (PSM), and their candid perceptions about the presence of each of the system action situation and organization action situation conditions in the qualitative research. I also conducted semi-structured interviews with 8 fellows from NAPA who did not take a survey. Responses from the NAPA fellows, which included former policy makers and senior federal officials, some of whom personally participated in the legislation and implementation of GPRA, PART, and GPRAMA, provided primary evidence used in chapter 3 on the history of performance management reform in the U.S. Government. Interviews were conducted from March to April of 2021 using Virginia Tech Zoom teleconferencing system, en lieu of in-person interviews, to prevent the possibility of transmitting COVID-19.

Data Collection and Preparation Procedures

In the fall of 2020, I developed the online non-sampling survey instrument following guidelines in Dillman, Smyth, and Christian (2009) and recommendations of my committee members. I used closed-ended questions with "yes/no/don't know" response options to determine whether or not, in the expert judgment of the federal manager, each performance management practice had been institutionalized in the manager's agency. I used a five-point

[68] The data for the qualitative portion of this paper was generated using Qualtrics software, Version 032021 of Qualtrics. Copyright © 2020 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com

bipolar ordinal scale to measure the degree to which the respondent agreed or disagreed with a provided statement that a system action situation condition or an organization action situation condition was present in their agency. The questions were framed to concisely describe conditions widely postulated in the public management literature (see chapter 2) as being significant predictors of performance management institutionalization, and were clustered together by their logical alignment to inferential model conditions (excluding the latent variable of agency identity, which was measured using 2017 GAO survey data only).

After drafting the survey, I subjected it to three rounds of review by fellow PhD students in the Center for Public Administration and Policy and by several graduate school faculty members. Following best practices of other effective surveys, such as a recent Director of National Intelligence (DNI) survey on workforce responses to the COVID pandemic (link is no longer active), I was attentive to the importance of providing survey participants unambiguous definitions of key concepts. After approval of the survey and interview questions by my committee chair and the IRB, I began to recruit volunteers to participate in the survey and interviews. The final survey is at appendix H. Appendix I is the recruiting flier. Appendix J presents the email templates I used to communicate first with agency PIOs to whom I was referred by the OMB and GSA, and then with agency performance system manager volunteers who responded to the recruiting flier I emailed to the PIOs and DPIOs. Appendix K presents the consent form that was provided to all volunteers for them to review and respond with affirmative consent via email per IRB approval before they received a link to the online survey.

Recruiting busy federal officials to voluntarily participate in academic research is no easy task. They already have government-directed surveys they are asked to take, such as OPM's

Federal Employee Viewpoint/Best Places to Work survey[69] annually and the GAO's "managing for results" surveys every four years or so.[70] Senior agency officials must expend considerable leadership capital on a recurring basis to convince survey-weary agency employees to once again take these surveys and get agency response rates up. These public leaders therefore do not have the time or capacity to respond to every non-governmental survey or interview request they receive each year, even a survey that takes just 15 minutes or less. Some PIOs were upfront with me that mine was not among the favored few they could support in 2021. Understandably, they have little to gain, and potentially much to lose, if their agencies are reflected poorly in the final study. On the other hand, they have little to lose by not being represented in an academic study, if they expect the study will gain little attention or notice, or perhaps even if it does. Additionally, a poorly managed and executed survey could provide invalid results that could cast an agency in an unfair and invalid light.

To win over skeptical public leaders, researchers first need to gain access to very senior officials at higher levels, persuade them of the value and importance of their research for public service practice, convince them the research has some level of merit and credibility among experts and respected authorities already, and win their trust that the study results will not somehow be used against them or their organization. For this study, I am confident my association with the National Academic of Public Administration from almost the beginning added a sense of legitimacy to the effort and probably led to a higher response rate to the survey and interview invitations.

---

[69] https://www.opm.gov/fevs/reports/governmentwide-reports/governmentwide-management-report/governmentwide-report/2020/2020-governmentwide-management-report.pdf; https://bestplacestowork.org/rankings/overall/large

[70] https://www.gao.gov/products/gao-18-609sp

Most federal leaders would rather not have to play the "blame game" at all if they do not have to. Hood (2011) refers to this attitude as "anticipative blame avoidance" (6). Declining an invitation to join an academic study can be seen as a way to avoid potential blame for potentially unflattering or unjustified findings, or to save time and resources for higher priorities. The former concern may have played a role in some agencies' decision to not participate, but the latter concern was definitely a factor during this study, because much of the federal workforce was stretched thin in performing their duties due to COVID-related manpower shortfalls, extended teleworking arrangements, and finally, a change in Administrations in early 2021 that took place just as the study recruiting process was getting underway. The arrival of the new Administration in January 2021 created significant turnover among politically-appointed Department Heads, COOs, PIOs and even some DPIOs, who are normally career officials. Some agencies admitted they could not participate in the study due to the need to focus their efforts on supporting the presidential transition. All of this was, of course, completely understandable.

If potential participants perceive there is no threat of potential harm to agency legitimacy or anyone's reputation, and are assured that those superior to them in the government hierarchy have no objections to their participation, then agency leaders and managers are less likely to object if they do happen to have the time to participate. Members of my committee were instrumental to my success in reaching the right people and framing my project in a way that would reassure them of its legitimacy and credibility. For example, Dr. Patrick Roberts recommended I create a website to inform prospective volunteers about the study and myself, which I did and included a link to it in the recruiting flier.[71] In addition, one of my committee members, Dr. (Brigadier General, US Army, retired) Michael Meese, introduced me to one of his

[71] https://hoya86.wixsite.com/research

professional colleagues, the President of NAPA, Ms. Terry Gerton, and she scheduled a meeting for me and several NAPA fellows with extensive backgrounds and professional connections in the federal-level performance management community of practice. One of these fellows was John Kamensky, one of the most recognized and widely cited national authorities on public sector performance management reform and good government. He showed a sincere interest in my research and turned out to be not only a treasure trove of primary information and authentic historical perspectives, but also a well-connected Washington insider who mentored me generously and introduced me to a wide range of current and former senior officials, now NAPA fellows, who were instrumental during the whole history of the modern performance management reform movement. These included current senior executives in the Office of Management and Budget (OMB) and the Performance Improvement Council (PIC).

Via the good offices of NAPA, and per the research plan as approved by the Virginia Tech Institutional Review Board, I reached out to senior officials in OMB and the GSA (which administratively manages the PIC) to ask for their help to socialize my study with the PIOs and DPIOs at all 24 CFO Act agencies. OMB sent an email announcement to all PIOs that they would be receiving a recruiting flier from Virginia Tech shortly, and shared the names and official email addresses of the PIOs and DPIOs with me. I then sent the PIOs and DPIOs the recruiting flier using the approved email. Out of the 24 CFO Act agencies, 12 responded positively and decided to participate. Therefore, only organizational conditions in these 12 agencies were analyzed, potentially limiting the generalizability of findings.

I kept the survey open for six weeks, from February 4, 2021 to March 16th, 2021. After sending the voluntary consent information to 31 employees who responded to the recruiting flier with a statement of interest in participating, I followed up via email with 19 who did not provide

a response one or two times over the 6-week period to inquire about their consent and interest in proceeding. Eight employees gave consent and took the survey without requiring any follow-up from me. Of the 11 who received a single follow up, nine sent me their consent, and of these nine, seven took the survey. One of these seven failed to complete most of the survey and the responses provided were therefore not used in the study. For five out of 8 employees receiving two follow-ups, the follow-ups were successful in gaining their consent, and all of the consenting employees took the survey. Five employees who sent their consent never took the survey, so 77 percent of consenting recruits ultimately took the survey. I found that following up with interested candidates was essential to ensure consent was received from a sufficient sample. Appendix L provides summary statistics of the survey results.

I also invited ten employees who completed the survey and worked at eight pre-designated agencies to participate in a 1-hour online recorded and transcribed interview in order to clarify their responses to the survey questions, obtain examples of agency mechanisms and procedures for institutionalizing GPRAMA, and better understand their attitude towards performance management and their PSM. The original research plan had called for interviewing performance system managers in at least two agencies belonging to each of the four types of government bureaucracies as classified by the taxonomy of bureaucratic types proposed by James Q. Wilson in 1989, in order to incorporate the interview data into the analysis for research question two, presented in chapter 5. Six performance management system managers from five (SBA, DHS, HUD, NASA, and Treasury) of the eight targeted agencies for interviews accepted and were interviewed. The Social Security Administration, Department of Energy, and Department of Education were the other three, and all declined to participate in the study.

Fortunately, at least one agency from each Wilson bureaucratic type was represented, a fortunate outcome. Appendix M presents the interview questions for the federal performance managers.

Finally, with NAPA's endorsement, John Kamensky introduced me to 11 other fellows, eight of whom I interviewed online after I received their consent via email. Appendix N presents the interview questions for the NAPA fellows. Their statements added valuable insights for chapter 3 on the history of performance management reforms in the federal government which were available nowhere else that the author has found. All instruments, consent information, and email templates received Virginia Tech Institutional Review Board approval on December 8, 2020 (IRB #20-814).

Data preparation was less complex than that employed in the quantitative analysis to comply with standard regression modeling requirements, but I followed certain steps to facilitate analysis of the data and interpretation of findings. To clean the 2021 survey data, I downloaded the raw response dataset from Qualtrics as a spreadsheet, eliminated superfluous data fields added by the survey software, examined each observation (participant responses to all questions) for completeness, converted "Yes" and "No" responses to 1 and 0, and converted the plain text responses to their respective ordinal values from 1 to 5. "Don't know" responses were left blank and not included in calculations of either outcomes or conditions. I also added fields for the inferential model outcomes and conditions based on the survey structure to the qualitative dataset. For the two institutionalization vectors (PM and PI), I created measures of positive agency outcomes that added up the number of "Yes" (or 1) responses and provided a scaled index of institutionalization that implemented the operationalization strategy described in chapter 2. To measure the presence of the system action situation condition of politics and resource dependency, and the organization action situation conditions, the survey used ordinal scale

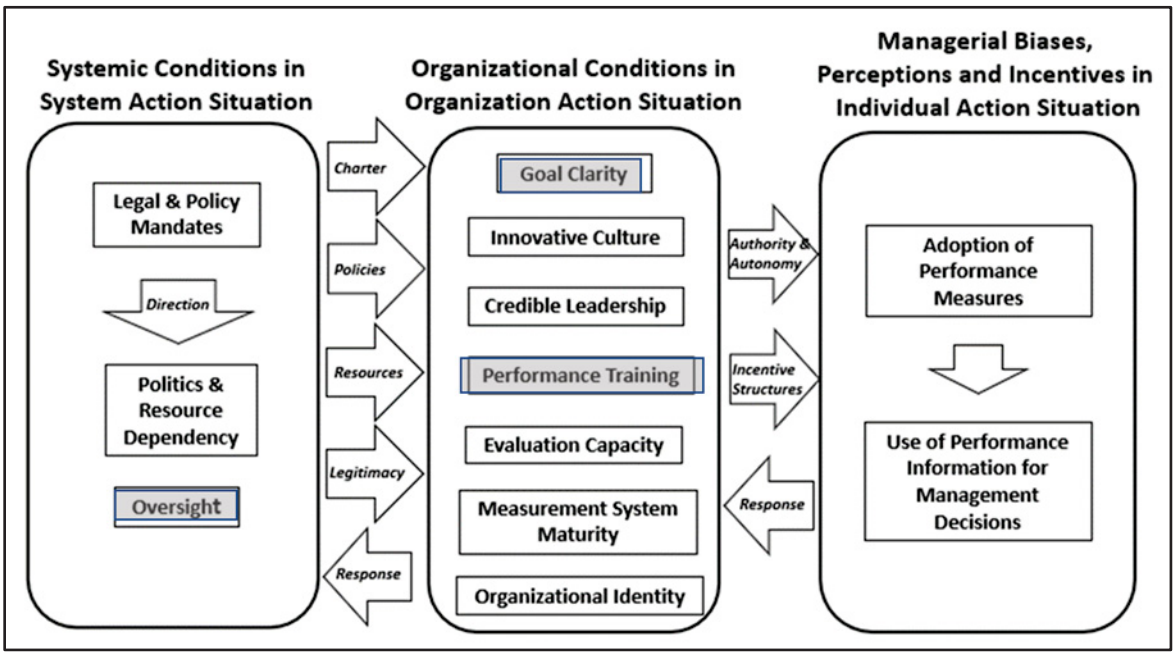
responses to calculate values based on the number of responses to survey questions associated with each condition that scored "somewhat agree" (4) or "strongly agree" (5). I selected level 4 as the threshold for assuming that the respondent perceived a sufficient presence of the condition in his or her agency. This aligned the standard used in the quantitative research with the ordinal threshold used in the partial proportional odds model in which manager perceptions rise above ambiguity to a degree of certainty and salience to the individual manager.

After first explaining the qualitative analysis strategy and introducing QCA, the next section describes the survey responses and the literature-informed method used to calibrate the agency-level aggregated survey data for the outcomes and conditions for the QCA analysis. Audio and video of all of the interviews with federal performance system managers and NAPA fellows was recorded by Zoom, which automatically generated rough electronic transcripts for archival purposes and to share with participants who requested them.

Construction of Conditions

In the quantitative analysis using base models 1 and 2, the shaded system action situation and organization action situation variables in the inferential model (figure 4.14) were found to be statistically or substantively insignificant predictors for GPRAMA institutionalization at the governmentwide level. The unshaded variables in the organization action situation were significant and were used in the qualitative research portion of the study.

Figure 4.14. Inferential Model Conditions for the Qualitative Analysis



Source: Author's own design.

Political and resource dependency influences were not assessed in the quantitative models due to lack of suitable GAO survey data, but were included in the qualitative research design. External oversight and goal clarity were not significant for either index of institutionalization and were associated with only two (goal clarity) or three (oversight) performance information uses. For oversight, these two uses were arguably explained by the role of parallel federal institutions for contracting and budget oversight more than performance management, so positive manager perceptual associations should be expected. For goal clarity, only its association with performance goal setting was significant at the .05 level, and managers setting performance goals, almost by definition, generally try to set clear ones, so additional efforts to investigate the role of goal clarity would not be especially fruitful for discovering new relationships at the agency or micro levels in this study. Performance management training was significant for the adoption of performance measures, but confusingly not for the adoption of any

specific kind of performance measure. For this reason, along with the fact that its regression coefficient was the lowest of all predictors, I judged it to be insufficiently relevant for the qualitative research phase.

Several conditions, however, made it through the correlational screening. As shown in table 4.9, an accountable and empowered culture was strongly and substantively significant for all institutionalization indexes. Credible leadership commitment was significant only for the adoption of efficiency measures and for the use of performance information for six of nine management uses of performance information, and not significant for the adoption of performance measures or use of performance information generally. Therefore, only its relevance for the use of performance information was included in the qualitative research. Like culture, capacity for evaluation of performance, the first control variable, was as strongly and substantively associated with all indexes of institutionalization as culture was, and was retained in the qualitative phase. Managers perceived the maturity of their agency's performance measurement system as positively associated with the adoption of performance measures. But the association does not clearly reveal the direction of the relationship, and it is certainly possible that the maturity of the measurement system may be a result of the adoption of a range of performance measures; logically it would seem so. So, there is an unanswered question for this variable that lends itself to further qualitative research to clarify. Therefore, this variable was retained as a provisional qualitative condition for the adoption of performance measures. How PSMs view the identity of their organization cannot be clearly ascertained from quantitative measures alone, so discovering how members of an agency perceive its mission-unique character is also included in the qualitative research plan.

For the individual action situation in figure 4.14, the two associated conceptual blocks collectively represent manager decisions and actions to institutionalize GPRAMA in their organization. The study concedes that there are likely dozens of other decisions agency managers must make that are necessary for the institutionalization of GPRAMA, but these represent the two primary categories of institutionalization indicators proposed to date in the performance management literature. The area surrounding these blocks in the model represents mediating biases, attitudes/perceptions, and incentives affecting managers' decisions to institutionalize the law into routine practices. The latter conditions were not quantitatively modeled in phase 1, but are included in the multiple case study analysis to ensure study findings explicate the complete inferential framework to the greatest degree possible commensurate with this study's data collection capability, which is admittedly limited relative to the research scope.

Success for all positivist research designs demands a high level of precision in defining and operationalizing variables, outcomes, and conditions. QCA procedures require researchers to deeply engage with theories and case evidence related to model conditions. Chapter 2 provided a comprehensive literature review of the conditions in the inferential model and presented the sometimes-contradictory theoretical claims about their influence on the institutionalization of performance management practices. Just as I did for the latent variables in the quantitative models, I created measurable institutionalization outcomes and system action situation and organization action situation conditions for the QCA and case study analyses. While the outcomes used the same indicators from the quantitative models (plus "other," if applicable), the conditions combined multiple theoretically-informed indicators noted in the literature review which expanded on the indicators used for the quantitative variables which were limited by the GAO survey. This means that the statistically-significant model predictor

variables from the quantitative analysis are specified more completely as conditions in the qualitative phase by use of a more comprehensive set of literature-informed indicators. The survey and federal manager interviews incorporated questions which were indicators of politics and resource dependency as a system action situation condition and indicators of the significant organization action situation conditions.

The qualitative analysis relies on observational data and does not estimate latent variables as the quantitative phase did. For the two kinds of GPRAMA institutionalization outcomes, PM and PI, I calculated the mean of manager "yes" responses by respondent and averaged respondent index scores clustered by respondents' agency affiliation. Higher scores for outcomes imply performance system managers perceive more kinds of associated institutionalized practices in their agency. Lower scores imply performance system managers perceive the absence of associated practices. As noted in the previous section, for each observation in the survey dataset, condition scores were the sum of all responses to condition-associated survey questions scoring a 4 "somewhat agree" or a 5 "strongly agree." Agency-level condition scores were the mean condition scores for all observations of managers from that agency responding to the survey. The outcomes and conditions were thus converted from dichotomous nominal (outcomes) and Likert-scale ordinal (conditions) data at the individual level of observation in the 2021 survey to ratio-level data at the agency level of observation. This conversion made the survey data suitable for calibration into fuzzy set data. The values in table 4.13 served as the raw data for the QCA analysis. These are agency-level index averages of individual respondent outcome and condition scores. Individual observations are not shown, nor are the number of observations by agency, in order to protect participant anonymity.

Table 4.13. Agency Average Scores from 2021 Federal Performance System Manager Survey

	Institutionalization Outcomes		System and Organizational Conditions				
	PM	**PI**	**P&R**	**C**	**L**	**X**	**M**
USAID	6.00	8.00	5.50	4.00	6.00	4.50	7.50
DOC	6.00	6.00	5.00	4.00	1.00	1.00	6.00
DOD	6.00	6.00	6.00	3.00	6.00	5.00	7.00
HHS[72]	5.00		4.00	3.00	5.00	5.00	5.00
DHS	4.40	4.40	2.80	4.60	4.40	4.40	7.00
HUD	6.00	8.00	3.00	2.00	1.00	4.00	6.00
DOJ	3.00	4.67	4.00	4.00	4.00	3.67	6.67
DOL	3.00	6.00	3.00	5.00	4.00	3.00	8.00
NASA	4.00	4.50	2.50	6.50	6.00	3.50	5.50
SBA	5.00	6.00	2.00	4.00	7.00	4.00	8.00
DOT	3.00	5.00	3.00	6.00	6.00	3.00	7.00
Treasury	4.00	6.00	3.00	3.00	5.00	1.00	5.00
MAXIMUM	6.00	8.00	6.00	6.50	7.00	5.00	8.00
MINIMUM	3.00	4.40	2.00	2.00	1.00	1.00	5.00
MEAN	4.62	5.87	3.65	4.09	4.62	3.50	6.55

Notes:
PM Adoption of Performance Measure
PI Use of Performance Information
P&R Politics and Resource Dependency
C Accountable and Empowered Culture
L Credible Leadership Commitment
X Capacity for Evaluation of Performance
M Measurement System Maturity

Units of Observation and Analysis and Measurement Models

In the qualitative research, the units of observation are federal performance system managers (PSM), who are members of agency performance management teams (PMT) and voluntarily answered questions in an online survey and during interviews. Observed data are

[72] HHS participants provided no responses to questions about the uses of performance information (PI). As they were not among the agencies interviewed after the survey, their reasons for not responding to those questions are unknown, and that is missing data. Like most statistical software programs, FSQCA skips observations with missing data using listwise deletion as a default technique.

aggregated at the agency level to enable cross-agency comparisons and to protect participant anonymity. The unit of analysis is the program of GPRAMA institutionalization administered by the headquarters-level PMT in each participating agency. PMTs typically consist of the PIO, the Deputy PIO, and full-time government employees who are professionally trained and compensated for managing the institutionalization of performance management practices in their agency. The rationale for collecting perceptual data from federal performance system managers is that they are the government employees who are best positioned within their organizations to observe and make sense of the presence and influence of politics and resource dependency, the agency's culture, its leadership's commitment, agency capacity to evaluate performance, and the maturity of the agency's measurement system on the PMT's ability to institutionalize GPRAMA. They have daily opportunities to make such observations during meetings, presentations, training sessions with line managers and employees, feedback and guidance sessions with agency senior leadership, including the head of the department, the COO, and the senior executive leaders of the operating units.

A limitation and important caveat on the unit of analysis is that some agencies had more robust participation by members of their PMT than others did, which likely makes the cohort of study participants a better representative proxy for the PMT in those agencies. All agencies had in common that no PIOs, all of whom are political appointees, chose to participate. However, seven Deputy PIOs did, including four senior executives and three GS-15 level managers. In five of the 12 agencies, PMT participants only included mid-level (GS-13/GS-14) performance analysts, or upper-level specialists or division/team leaders in the headquarters performance management shops (GS-14 or GS-15). Employees at these non-executive levels generally had less opportunity to observe the effects of politics and resource dependency as DPIOs did; survey

and interview responses generally confirmed this. On the other hand, I expected that they would have an equal or better opportunity than PIOs and DPIOs to observe the effects of organizational conditions on the institutionalization process based on their interactions with business unit line and performance system managers. It is very probable the grade level, responsibilities, and professional perspectives of PMSs participating in the survey and interviews shaped how they responded to survey and interview questions. As a result, the data, findings, and interpretation of results in this study regarding agencies must be understood as fundamentally biased by the variable effects of work roles and organizational placement of participants, which were not uniform across the set of agencies studied, on their perspectives and attitudes towards GPRAMA institutionalization.

As outlined in the introduction to this chapter, the qualitative analysis is organized into two sub-studies using different analytic methods and models. First, a meso-level sub-study using survey data and QCA methods is used to identify necessary and sufficient causal configurations of significant system action situation and organization action situation conditions that help to explain the institutionalization of the two studied GPRAMA practices in the participating agencies. Then, a sub-study using a comparative multiple case studies design is employed to triangulate results of interviews with federal performance system managers (PSMs) with survey and QCA findings and to discover empirical evidence that characterizes the conditions in the context of participating federal agencies. Federal performance system manager attitudes and biases concerning performance management and the institutionalization of GPRAMA in their agencies discovered during interviews is also incorporated into the case study findings. The following description of the analytic strategy for each sub-study is presented accordingly.

Qualitative Sub-Study 1: Causal Configurations for GPRAMA Institutionalization

As described previously, there were several limitations of the quantitative research in this study. An additional and more fundamental limitation can be seen in the very nature of regression models. Rihoux et al. (2009) state "…regression-based methods focus primarily on the problem of estimating the net, independent effect of each variable included in an analysis on the outcome" (171) which is what the quantitative phase of this research did in order to identify which system action situation and organization action situation conditions are significantly and substantially related to the dependent variables. The quantitative research pursued a variable-oriented strategy, however, and such strategies have inherent limitations. Specifically, Ragin (2014) argues that ". . . a variable-oriented strategy is incapacitated by complex, conjunctural causal arguments requiring the estimation of the effects of a large number of interaction terms or the division of a sample into many separate subsamples" (69). To discover the configuration of potentially complex causal paths leading to institutionalization and non-institutionalization of GPRAMA practices, use of a quantitative approach was impractical for these reasons. I employed a meso-level (cross-agency) structured QCA of GPRAMA institutionalization cases using the fuzzy-set QCA (fsQCA) method to study the raw data in figure 4.19.

In this study, a causal path is a configuration of conditions whose presence or absence results in GPRAMA institutionalization in an agency. To answer research question 1, I sought to discover which conditions in the inferential model are present or absent when agencies are successful in institutionalizing GPRAMA practices. Also, which are present, or more likely absent, when agencies fail to institutionalize it? Are any conditions necessary for any institutionalization to occur? Are any sufficient by themselves or, more likely, in combination with others? Based on the assumption of causal asymmetry, there is no reason to expect that the absence of the causes that brought about a positive outcome should explain the non-occurrence

of the same outcome (Ragin 2008 as cited by Fiss 2011, 394). Outcomes and non-outcomes are likely to have their own causal configurations. The fsQCA method can be used to answer these questions, and can help to answer the first research question, develop governance theory, and improve public management strategies to implement reform policies.

All variants of QCA are well-established research tools that support comparative case analysis, but they are by no means universally endorsed or relevant for all research contexts and purposes (Gerrits and Verweij 2016, 13).[73] It has been in use by scholars in many fields of social science since the late 1980s and early 1990s, originally by political scientists and historical sociologists (Berg-Schlosser et al. 2009, 3). Since then, it has been embraced by scholars in other fields and disciplines, including business, criminology, economics, geography, the life sciences, management, organizational behavior, and psychology (De Villiers 2017, 2; Yamasaki and Rihoux 2009, 123-124). Greckhamer et al. (2018) observe a "resurgence" in management research using configurational methods employing set-based analysis (483).

In public policy research, QCA has been receiving increased attention because it supports cross-case analysis of small to intermediate numbers of cases, quite common in public policy scholarship (Fischer and Maggetti 2016, 2) Public administration research using QCA began about 2012 and has demonstrated analytic merit for studying decision making in public organizations and the contributions of public employees to agency implementation of policies

[73] For example, while there have been recent efforts to enable QCA to support longitudinal research, Gerrits and Verweij (2016) note ". . . its essentially a-temporal nature poses challenges for using QCA to study complex systems and phenomena. . ." (13). Nevertheless, temporality can be integrated into QCA research via qualitative within-case investigation of longitudinal factors, or by adding a temporal dimension to the conditions or to the definition of the cases (De Meur, Rihoux, and Yamasaki 2009, 161-162). Other major methodological controversies center on errors in measurement and the effects of calibration decisions on the robustness of QCA models (M. Fischer and Maggetti 2017, 3), as well as the recent claims of possible bias in QCA Boolean optimization algorithms borrowed from the engineering sciences (Thiem et al. 2020).

(Thomann and Ege 2020). QCA-based scholarship has compared the institutionalization of public policy reforms at different levels of analysis. Examples include variance in antecedents of public pension reforms between nations (e.g. Carrera and Angelaki 2020), education policy reforms at the individual level in the context of universities (e.g. Bingham, Dean, and Castillo 2019) or at the state level for elementary and secondary schools (e.g. Freitag and Schlicht 2009), and asylum policies at the state level (Sager and Thomann 2017). Despite diverse searches through the QCA community's COMPASSS.org website and in other source databases, the author has found no examples of published empirical research using QCA to analyze the institutionalization of performance management reforms in the public sector,[74] so the present research is likely to be an original empirical contribution to QCA-based scholarship.

Because QCA is a relative newcomer to the fields of public policy, public management research, and public administration, a brief overview of the method is in order. QCA applies the comparative case method to identify similarities and differences between cases when the number of cases is relatively small (Ragin 2014, xxi, 12), but large enough to group case characteristics by variable and compare them across cases (Yin 2014, 174). Among other uses, researchers can employ it to develop new theoretical claims, test conjectures, and test hypotheses or existing theories (Berg-Schlosser et al. 2009, 15). According to Blatter (2017), QCA stands on strong elementarist ontological presuppositions with a degree of holism. He claims it has a truth-seeking epistemological purpose that includes a measure of sense-making. thereby situating it

[74] Most research on the institutionalization of performance management reforms the author could identify tends to use single case studies (Ho 2007; Ohemeng 2011), longitudinal case studies (e.g. Yetano 2013), or quantitative approaches (e.g. Nam 2019). I found a single unpublished manuscript presented at the January 2020 Public Management Research Association Conference that uses fsQCA to model the institutionalization of performance management in the context of Los Angeles (see Jackman, Musso, and Weare Undated) and am grateful to the authors for their written permission to cite it here.

conceptually between the analytic strategies of the comparable cases strategy, and causal process tracing (3). For a clearer visualization of its relationship with other forms of qualitative research, Appendix O shows the placement of QCA in the spectrum of ideal-type classes of qualitative methods.

As a way of analyzing differences between cases holistically, QCA provides a method to identify configurations of theoretically-informed conditions that are sufficient for an outcome to occur (Berg-Schlosser et al. 2009, 6). QCA creates a gateway for further research using single case and comparative within-case comparisons to understand how causal mechanisms lead to an outcome and are associated with conditions (Schneider and Wagemann 2013, 306-310). This study walks through that gateway and interprets the results of interviews as within-case evidence of agency practices for GPRAMA institutionalization and their relationships to agency-unique conditions in the causal configurations.

QCA extends the long traditions of case study research in the social sciences and is designed to help scholars investigate research questions concerning "…macrosocial units with meaningful social identities" (Ragin 2014, 16-17). Berg-Schlosser et al. (2009) note QCA methods are employed by researchers studying "meso-level" units of analysis such as "…organizations, social networks, collective actors, etc." (4). De Villiers (2017) argues QCA promises improved predictive validity over multiple regression-based models (1-2). Thomann and Maggetti (2020) claim the authentic case knowledge generated using a QCA approach in a research design enhances internal validity (3). Thus, I chose QCA, specifically the fuzzy set type, as a research strategy to answer the research questions using qualitative methods.

In all types of QCA, dependent variables are dubbed "outcomes" and independent variables "conditions."[75] These titles remind users that QCA is fundamentally a tool for case-centric analysis grounded in context, not a variable-centric method (Ragin 2008, 4-5) . In fuzzy set QCA (fsQCA), cases are defined as configurations of values for the degree of membership of the unit of analysis (here, federal agencies) in theoretically-informed condition sets which, when combined in different ways, create causal paths to outcomes (here, the institutionalization of GPRAMA practices). In the event of "multiple conjunctural causation," different causal paths may also even lead to the same outcome (Berg-Schlosser et al. 2009, 6-8). This implies that if a government agency as a unit of analysis has equifinal causal paths to the same outcome (e.g., institutionalization of a GRAMA practice), then there may be more than one case associated with that agency, and it is certainly possible other agencies may share that causal configuration. This approach is well-suited for this study, because the analytic question of interest is which combinations of system and organization action conditions must be present or absent to lead to the institutionalization of GPRAMA, and which combinations lead to non-institutionalization?

Coding of outcomes and conditions using fsQCA is based on the empirical case data that is collected and then calibrated into fuzzy sets with values between one and zero, whereby one means the unit of analysis is a full member of the condition set, zero means the unit of analysis is a nonmember, above 0.5 implies the unit of analysis is more in than out of the set, and below 0.5 means the unit of analysis is more out than in (Ragin 2008, 30). QCA methods use a "truth

[75] "Outcomes" are self-explanatory; they are results. "Conditions" refer to either so-called INUS conditions, which means "insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result" or to SUIN conditions, which mean "sufficient but unnecessary part of a factor that is insufficient, but necessary for the result" Causal configurations are combinations of INUS and SUIN conditions. (Mackie 1974 as cited by Schneider and Wagemann 2013, 79-80).

table" to identify and analyze all possible causal configurations that lead to a common outcome (Ragin 2008, 23-24). The number of possible causal configurations (cases) in truth table rows is 2^k where k = number of conditions (Schneider and Wagemann 2013, 334). In fsQCA, cases with membership scores of less than .5 in the outcome set are considered to be logical remainders, or cases with insufficient evidence (Schneider and Wagemann 2013, 329). The use of Boolean logic by the Quine-McCluskey algorithm (McCluskey 1956) reduces causal complexity in the truth table by minimizing multiple causal combinations via automated means into more parsimonious models of causality that retain only those conditions that are either necessary or sufficient, or both (Rihoux and De Meur 2009; Klir et al 1997 as cited by Schneider and Wagemann 2013, 104). Berg-Schlosser and De Meur (2009, 28) suggest it is common practice for an intermediate-N (10-40 cases) QCA study to use four to six or seven conditions. This study conforms with these practices by using five conditions for 12 cases.

The ultimate goal for using QCA methods is to implement a formalized analytical strategy to compare cases in order to overcome subjective interpretation of data and results and achieve "modest" generalizability of findings, while enabling other scholars to replicate the study (Berg-Schlosser et al. 2009, 11-14). There are three basic types of QCA currently in use; a fourth type, generalized set QCA (gsQCA), exists but had not been used in published research by 2017 (Thiem 2017, 426). Each of the core types employs different set-theoretic analytic approaches: crisp-set (cs) QCA using binary sets; multi-value (mv) QCA using nominal or ordinal sets; and fuzzy-set (fs) QCA, using continuous sets, with thresholds for calibration, that express the degree of membership of the unit of analysis in the set. All three core QCA types are well-suited to help researchers identify complex causal pathways because they are sensitive to contexts of the units of analysis and observation and are designed to 1) compare a similar set of

cases that respond differently to the same conditions, 2) study contexts which are situated between grand social theories and individual-level or single case studies, such as government agencies, and 3) identify conditions, and conjunctural combinations of conditions, which are sufficient for an outcome to occur (Berg-Schlosser et al. 2009, 6-8). All forms of QCA are suitable for analysis of intermediate-N research of between 10-100 cases as well as large-N research contexts of more than 100 cases (Rihoux et al. 2009, 169). Method selection depends on the number of cases and "…the necessity to preserve the richness of the data information in the raw data set" (Herrman and Cronquist (2008) as cited by Rihoux et al. 2009, 169).

The choice of appropriate QCA technique depends on a trade-off choice between the goals of simplifying causal complexity and preserving empirical case information. For example, where rich case data is less important for achieving the research goal than parsimonious model results are, the use of csQCA offers the benefit of causal and relative procedural simplification with less risk to analytic objectives from the acknowledged loss of data that occurs via the simplifying process of condition dichotomization (De Meur, Rihoux, and Yamasaki 2009, 149). In fact, binary or categorical variables are often employed in institutional analyses involving multiple cases, but a few scholars have developed methodological approaches to support research designed to identify antecedents of policy institutionalization that operationalize variables based on their magnitude (Baldwin, Chen, and Cole 2019, 907). The latter approach is appropriate when differences between cases are a matter of degree, and not of kind (De Meur, Rihoux, and Yamasaki 2009, 149). The model of policy institutionalization proposed by Tolbert and Zucker (1996) described in chapter 2, a phased process leading to sedimented practices, suggests institutionalization is best measured on a continuum of embeddedness, and not as a binary or categorical outcome. Depending on the stage of reform implementation they are in,

federal agencies can be full, partial, or nonmembers in the sets of agencies with sedimented practices of performance measure adoption and performance information use. The same is true for their membership in the antecedent conditions. Because of this, fsQCA is the most appropriate QCA technique for this study.

The section on the construction of conditions identified the conditions that were deemed to be statistically and substantially significant in the quantitative modeling for each of the GPRAMA institutionalization outcomes. Culture (C), Capacity (X), and Maturity (M) were significant for Performance Measure Adoption (PM), and C and X were significant for Use of Performance Information (PI). In addition, because Credible Leadership Commitment (L) was positively associated with 6 of 9 specific uses of performance information, it is treated here as a significant predictor for PI. The outcomes and conditions were re-operationalized as conditions for the qualitative analysis.

Table 4.13 presented the raw, agency-level data for each of the outcome and condition measures for the fsQCA analysis. Before they can be used in fsQCA, research measures (i.e. outcomes and conditions) must be calibrated into fuzzy sets between 0 (full non-membership) and 1(full membership) using qualitative anchors to demarcate thresholds for full membership, non-membership, and the point of maximum ambiguity as to membership status (the cross-over point); investigators must also stipulate how membership is graded between the anchors (Schneider and Wagemann 2013, 32). Calibration is common in the hard sciences, but less common in the social sciences, where research commonly uses uncalibrated, and thus inferior, measures to compare cases with each other in a relative way, not against fixed standards (Ragin 2008, 72). Justifying the basis for anchors as thresholds is good calibration practice and adds to transparency and replicability in research, which in turn adds legitimacy and creates

opportunities for honest intellectual dialogue on cases, contexts, and conceptual frameworks (Schneider and Wagemann 2013; De Meur, Rihoux, and Yamasaki 2009, 151).

Calibration thresholds should be based on external standards which are either theoretical or empirical (Schneider and Wagemann 2013, 32). Standard thresholds for recognizing levels of institutionalization of performance management practices in the public sector, or for helping to measure the degree of agency membership in the set of agencies with high levels of accountable and empowered cultures, credible leadership commitment, capacity for evaluation of performance, and mature measures do not appear in the literature. Even U.S. Government oversight organizations like the GAO, which closely monitor and report to Congress on agency implementation of GPRAMA on an ongoing basis and publish a series of evaluations that present agency institutionalization indexes based on perceptual surveys of federal managers, base their assessments on agency progress in implementation relative to other agencies, and to themselves over time, but not against pre-defined standards of attainment or any scheme of evolutionary classification. In other words, based on GAO data, a federal agency may be the best among federal agencies in terms of using performance information for management decisions, yet still not have performance information use practices that are fully sedimented and widely used in routine operations at all levels of the organization. Without calibrated data, there is no yardstick. Therefore, based on GAO data alone, it is impossible to infer the absolute level of institutionalization of GPRAMA practices in federal agencies.[76] How, then, did I calibrate the raw data? Following the example of Freitag and Schlicht (2009), endorsed by Schneider and

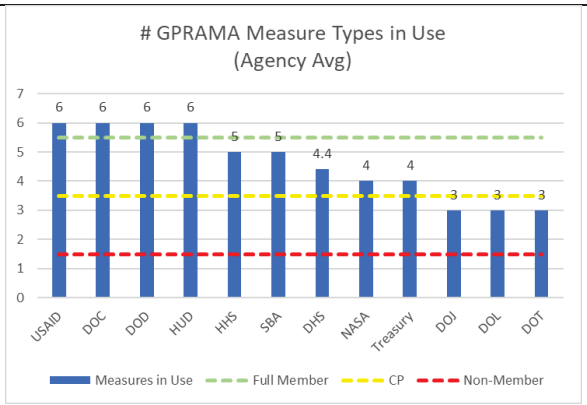
[76] In connection with associations built with GAO members from prior coordination during the author's doctoral studies, the author corresponded with GAO several times during this research and solicited their thoughts about latent thresholds that GAO had in mind when evaluating its survey data for compliance and implementation by agencies. GAO leaders considered the request but provided no response.

Wagemann (2013, 36-37), I selected thresholds for fully-in membership and fully-out membership based on clustering and large value breaks among case clusters, applying theory and empirical knowledge outside of the raw data wherever possible for each outcome and condition. Appendix P presents descriptive data for each outcome and condition set based on PSM survey responses, as well as the selected calibration thresholds and the justification for each threshold. Figures 4.15 and 4.16 show the calibration thresholds in graphical form for the two types of GPRAMA institutionalization outcomes and five conditions of the system and organization action situations.

Figure 4.15. Calibration of Outcome Sets for Fuzzy-Set QCA

Calibration of GPRAMA Institutionalization Outcomes

Adoption of Performance Measures (PM) **Use of Performance Information (PI)**



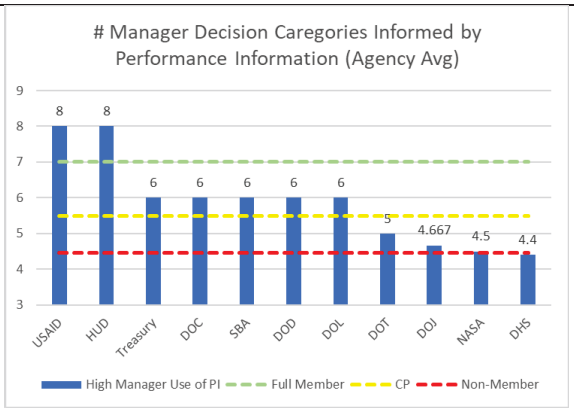
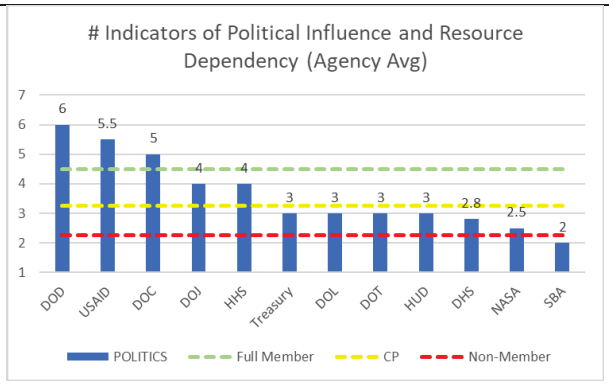
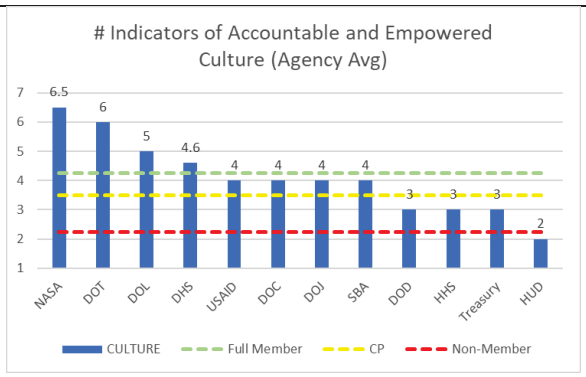
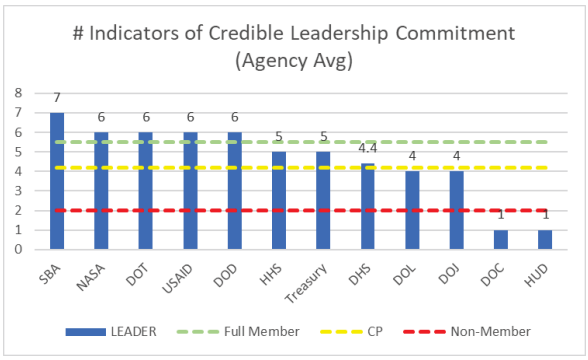


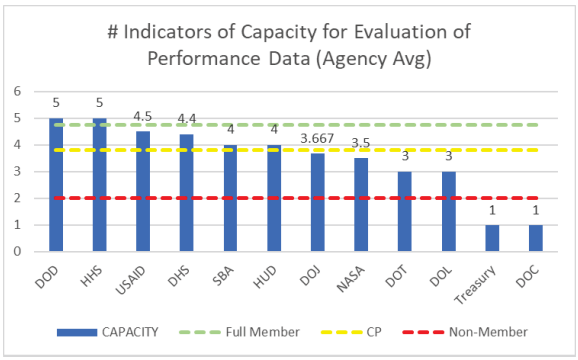
Figure 4.16. Calibration of Condition Sets for Fuzzy-Set QCA

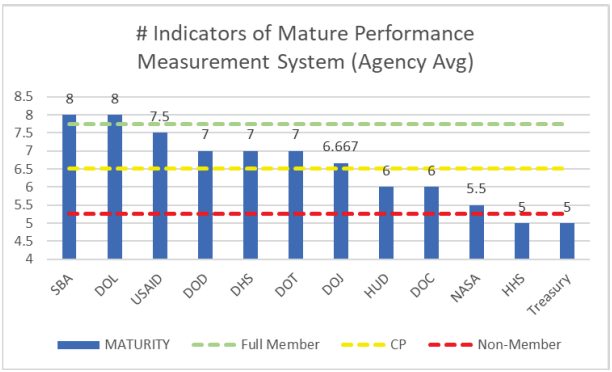
Calibration of GPRAMA Institutionalization Conditions











Using the FSQCA 3.0 software[77] and procedures in Ragin (2008), I calculated fuzzy-set values based on these calibrations. These are shown in table 4.14.

[77] The free software is downloadable at http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml

Table 4.14. Agency Fuzzy Set Data from 2021 Federal Performance System Manager Survey

	Institutionalization Outcomes		System and Organization Action Situation Conditions				
PSMs in Agency	**PM**	**PI**	**P&R**	**C**	**L**	**X**	**M**
USAID	0.98	0.990	1	0.88	0.98	0.9	0.92
DOC	0.98	0.730	0.99	0.88	0.01	0.01	0.23
DOD	0.98	0.730	1	0.18	0.98	0.98	0.77
DOJ	0.32	0.080	0.86	0.88	0.43	0.44	0.6
DOL	0.32	0.730	0.32	1	0.43	0.21	0.97
HHS	0.9		0.86	0.18	0.86	0.98	0.03
DHS	0.79	0.040	0.21	0.99	0.61	0.87	0.77
HUD	0.98	0.990	0.32	0.01	0.01	0.65	0.23
NASA	0.68	0.050	0.1	1	0.98	0.38	0.08
SBA	0.900	0.730	0.02	0.88	1	0.65	0.97
DOT	0.320	0.190	0.32	1	0.98	0.21	0.77
Treasury	0.680	0.730	0.32	0.18	0.86	0.01	0.03

Notes
P&R Politics and Resource Dependency
C Accountable and Empowered Culture
L Credible Leadership Commitment
X Capacity for Evaluation of Performance
M Measurement System Maturity
Missing data for HHS for PI

The fsQCA modeling results, analysis, and findings are presented in the section, "Qualitative Findings," later in this chapter.

Qualitative Sub-Study 2: Case Studies of Five Federal Agencies

To support the development of theory regarding causal antecedents of GPRAMA institutionalization, I conducted the quantitative research and qualitative sub-study 1 using variable-oriented research designs, which Ragin (2014) argues are "theory-centered" (53). As he observes, a basic goal of social science is to ". . . explain and interpret the diverse experiences of societies, nations, cultures, and other significant macrosocial units" (53). As anticipated, the evidence available for sub-study 1 using QCA methods was sufficient for its research goal of

achieving modest generalizability of findings, but, as with many cross-case studies, was wholly inadequate to fully characterize the "diverse experiences" of federal agencies in institutionalizing GPRAMA. Miles and Huberman (1994) note that in such studies ". . . the details of any specific case recede behind the broad patterns found across a wide variety of cases, and little explicit case-to-case comparison is done" in variable-centric research designs (174).

Any account of the story of GPRAMA institutionalization would not be complete without giving due respect to the facts on the ground and the perspectives of the federal performance system managers who are responsible for leading their agency's response to GPRAMA requirements. Yin (2014) argues that case studies, when used as an element of a mixed methods research design, may provide complementary evidence to "elucidate some underlying process" previously identified using quantitative or other methods, a strategy he defines as a "scenario of complementarity" in which ". . .the case study questions are likely to be closely coordinated with those of the other methods. . ." (193). In order to describe and compare the lived experiences of agency PSMs, qualitative sub-study 2 refocuses the research effort from the cross-case analysis in the QCA to develop middle-range theories about what conditions affect institutionalization at the meso level to a within-case analysis at the micro-level within federal agencies. In this sub-study, federal PSMs are thus the unit of observation and their agency's performance management system (PMS) is the unit of analysis. As noted previously, online interviews of six officials from five agencies were conducted to collect this data.

In this sub-study, the research goals were to complement the findings from the QCA by describing case-level evidence related to the postulated conditions in the causal configurations leading to positive and negative institutionalization outcomes, and explaining how the presence or absence of the conditions influenced the institutionalization of agency performance

management system structures and business practices in federal agencies. The cases are therefore partly descriptive and partly explanatory. After summarizing the descriptive findings from the survey and the QCA findings from sub study 1, sub study 2 develops within-case narratives related to the QCA findings for each agency using the survey and interview data. In addition, the five case studies present findings from the interviews regarding attitudes and biases of PSMs towards performance management and their agency's institutionalization of GPRAMA.

Qualitative Findings

Qualitative Sub-Study 1: Causal Configurations for GPRAMA Institutionalization

For qualitative sub-study 1, I conducted fsQCA modeling using the FSQCA v3.0 software and the agency-aggregated fuzzy set data for the outcomes and conditions in table 4.14. Table 4.15 provides descriptive statistics about the two institutionalization outcomes and the five conditions.

Table 4.15. Descriptive Statistics on Raw and Fuzzy Data for QCA Outcomes and Conditions

Raw Data	**N**	**Mean**	**STD**	**Min**	**Max**	**Missing**
PM	12	4.616	1.18	3	6	0
PI	11	5.869	1.18	4.4	8	1
P&R	12	3.65	1.21	2	6	0
C	12	4.092	1.24	2	6.5	0
L	12	4.616	1.83	1	7	0
X	12	3.505	1.29	1	5	0
M	12	6.555	1.01	5	8	0

Fuzzy Set Data	**N**	**Mean**	**STD**	**Min**	**Max**	**Missing**
PM	12	0.736	0.26	0.32	0.98	0
PI	11	0.544	0.36	0.04	0.99	1
P&R	12	0.527	0.36	0.02	1	0
C	12	0.672	0.38	0.01	1	0
L	12	0.677	0.36	0.01	1	0
X	12	0.524	0.35	0.2	0.98	0
M	12	0.53	0.37	0.03	0.97	0

Notes:
PM Outcome - Set of agencies with high level of adoption of performance measures
PI Outcome - Set of agencies with high level of use of performance information

P&R Condition - Politics and Resource Dependency
C Condition - Accountable and Empowered Culture
L Condition - Credible Leadership Commitment condition
X Condition - Capacity for Evaluation of Performance condition
M Condition - Measurement System Maturity condition
Missing data – PI is missing one agency for both tables due to no responses provided by survey participants in one agency related to agency use of performance information.

In model 1A of the fsQCA analysis, the outcome set is the set of agency performance management programs which have adopted a wide variety of performance measures, PM. Condition sets include only those determined as statistically significant in the quantitative research: the set of agencies with an accountable and empowered agency culture (C), the set of agencies with adequate capacity for evaluation of performance (X), and the set of agencies with a mature performance measurement system (M). The set of agencies which experience substantial political influences and resource dependency (P&R) is included as well. Model 1B is identical to model 1A, except the outcome set is negated (i.e. subtracted from one), in accordance with QCA best practice recommendations of Greckhamer et al. (2018, 490) and Schneider and Wagemann (2013, 279-280). This enables an analysis of the presence or absence of condition sets that may be necessary or sufficient to result in the non-adoption of performance measures. In model 2A, the outcome set is the set of agency performance management programs with high management use of performance information, PI, and the condition sets, again based on the statistically-significant variables, are C, L, and X. P&R is also included as a condition set. Model 2B is identical to model 2A, except the outcome set is negated in the same manner as in model 1B to identify causal configurations that result in the non-use of performance information.

With all outcome and condition sets well-defined based on theory and empirical contexts grounded in the literature, carefully selected based on results of the correlational analysis, and then populated with calibrated fuzzy data, the next step in the fsQCA modeling process as

recommended by Ragin (2009), before constructing a truth table,[78] is to check for the existence of any condition sets in the models which are "necessary," meaning that the outcome could not occur without it, but there is no assurance the outcome will occur if it is present as it may not by itself be sufficient (109-110). Only one condition set, the absence of C, had a level of consistency for being a necessary condition for the absence of the use of performance information (.81) that came anywhere near the recommended threshold of .90 for necessary conditions (Schneider and Wagemann 2013, 143). As it did not exceed the standard, it was retained for the analysis of potentially sufficient configurations.

In accordance with fsQCA practices, preparing a truth table requires the investigator to make certain assumptions. Because this is a smaller level intermediate-N analysis, I set the minimum number of cases for each causal configuration at 1, so the perceptions of all agency PMTs were included in the sufficiency analysis. If the case had a raw consistency value of .80 and a PRI consistency of 0.5 or higher as recommended by Greckhamer et al. (2018, 489), the outcome was coded as "1," otherwise zero. Figure 4.17 shows a screen shot of the truth table user interface in the FSQCA software. This example is for the truth table for model 1A.

Figure 4.17. Truth Table User Interface in the FSQCA Software

[78] A "truth table" is a matrix with columns that are sets (outcomes and conditions) and rows that show each possible AND combination of conditions based on their presence or their absence. The maximum number of rows in a truth table is 2^k, where k equals the number of conditions in the model (Schneider and Wagemann 2013, 92-93). In each configuration, or "mintern," a condition cannot appear more than once (Thiem 2017, 425).

FPOLITIC	FCULTURE	FCAPACIT	FMATURIT	number	FMEASH	cases	raw consist.	PRI consist.	SYM consist
0	1	1	1	2 (16%)	1	cases	1	1	1
0	1	0	1	2 (33%)	0	cases	0.661972	0.257732	0.257732
0	0	0	0	1 (41%)	1	cases	1	1	1
0	1	0	0	1 (50%)	1	cases	1	1	1
0	0	1	0	1 (58%)	1	cases	1	1	1
1	0	1	0	1 (66%)	1	cases	1	1	1
1	0	1	1	1 (75%)	1	cases	1	1	1
1	1	0	0	1 (83%)	1	cases	0.959799	0.910113	0.910112
1	1	1	1	1 (91%)	1	cases	0.947598	0.894737	0.894737
1	1	0	1	1 (100%)	1	cases	0.869565	0.54717	0.54717
1	0	0	0	0 (100%)		cases			
0	1	1	0	0 (100%)		cases			
1	1	1	0	0 (100%)		cases			
0	0	0	1	0 (100%)		cases			
1	0	0	1	0 (100%)		cases			
0	0	1	1	0 (100%)		cases			

The first four columns are the condition sets and the outcome set is "FMEASH." "Number" is the number of cases whose fuzzy set values exceed .5 for the configuration. The user can sort the table by any variable. Cases (causal configurations) with no empirical evidence are logical remainders and potential counterfactuals for the analysis.[79]

Using the Standard Analysis option, the FSQCA program simplifies the truth table to generate three alternative types of configurational solutions. Parsimonious solutions include all counterfactual configurations in the Boolean minimization, including those which are theoretically and empirically-supported, or "easy" counterfactuals, and those which may have empirical support but lack theoretical plausibility, known as "difficult" counterfactuals (Schneider and Wagemann 2013, 168). Parsimonious solutions minimize configurations to "core" causal conditions (Greckhamer et al. 2018, 490) but include all counterfactuals in the

[79] Ragin (2008) observes that counterfactual analysis has a long tradition in the social sciences going back to Max Weber, who referred to them as "thought experiments" (151). He notes analysis of hypothetical cases is prevalent in research using case studies, historical analysis, and social phenomena, where limited diversity is normal, and notes QCA treats counterfactual cases ". . . as substitutes for matched empirical cases" (150-151).

minimization regardless of their theoretical plausibility. Complex solutions include no counterfactuals and are thus based strictly on observed evidence. They typically generate more complex configurations with all possible core and contributing conditions due to the lack of counterfactuals which would improve causal simplification. Intermediate solutions include counterfactual configurations that are theoretically plausible and typically reveal both "core" and some "contributing" conditions based on the set of selected counterfactuals.

To select the counterfactuals to generate the intermediate solution, the investigator must define theoretical expectations for the presence or absence of each condition in each model based on existing theory and empirical evidence from cases concerning the conditions and their relationship to the outcome. These expectations are simplifying assumptions. FSQCA includes counterfactual cases conforming to these assumptions in the minimization process leading to the intermediate solution.[80] The researcher may choose to use any of these three types of solutions as the basis for comparative analysis. Ragin (2008) argues intermediate solutions achieve an interpretable and reasonable balance between parsimony and causal complexity and are especially useful when limited diversity is substantial (175). For these reasons, and because the data collected through the surveys and interviews provided adequate case-specific knowledge to

[80] The purpose behind expressing the theoretical expectation for the absence or presence of each condition in QCA models is to instruct the FSQCA program to select which logical remainders to retain as counterfactuals for the Boolean minimization process. Explicit acknowledgement of researcher assumptions concerning counterfactual cases is a strength of QCA methods vis-à-vis quantitative methods, which do not address the problem of limited diversity due to a simplistic approach to evaluating causation (Ragin 2008, 157-158) . Counterfactual cases which are implausible based on theory are deemed to be "difficult counterfactuals" and are eliminated by FSQCA prior to minimization for the intermediate solution to prevent unjustified simplification of the causal configurations as a result of leaving in highly unlikely counterfactuals (Ragin 2008, 163-166). Because Greckhamer et al. (2018) recommend justifying the selection of simplifying assumptions as a best practice for fsQCA (490), it is provided in appendix P.

justify the simplifying assumptions, I based my comparative analysis on the intermediate solution for all models.

For model 1A, for example, I set P&R as either present or absent based on the findings from the literature review and consistent with hypotheses H_{1a} and H_{1b} in chapter 2 which the historical evidence in chapter 3 supported. For the other three conditions, I based it on the preponderant views of scholars synthesized in the literature review and consistent with hypotheses H_4, H_5, H_7 and H_8. I expected the conditions to be present when the outcome of performance measure adoption was present, and to be absent when the outcome of performance measure adoption was negated in model 1B. I made similar theory-informed directional assumptions for models 2A and 2B. Tables 4.16 and 4.17 show the truth tables for all four models for the sufficiency analysis with consistency measures.

Table 4.16. Truth Table for Performance Measure Adoption/Non-Adoption (Models 1A and 1B)

					Outcome				
Mintern	**P&R**	**C**	**X**	**M**	**PM**	**n**	**Raw Consist.**	**PRI Consist.**	**Ideal-Type Agencies**
1	0	1	1	1	1	2	1	1	DHS, SBA
2	0	0	0	0	1	1	1	1	Treasury
3	0	1	0	0	1	1	1	1	NASA
4	0	0	1	0	1	1	1	1	HUD
5	1	0	1	0	1	1	1	1	HHS
6	1	0	1	1	1	1	1	1	DOD
7	1	1	0	0	1	1	.960	.910	DOC
8	1	1	1	1	1	1	.947	.895	USAID
9	1	1	0	1	1	1	.869	.547	DOJ
10	0	1	0	1	0	2	.662	.258	DOL, DOT
11	1	0	0	0	?	-			
12	0	1	1	0	?	-			
13	1	1	1	0	?	-			
14	0	0	0	1	?	-			
15	1	0	0	1	?	-			
16	0	0	1	1		?	-		

Table 4.17. Truth Table for Use/Non-Use of Performance Information (Models 2A and 2B)

					Outcome				
Mintern	**P&R**	**C**	**X**	**M**	**PI**	**n**	**Raw Consist.**	**PRI Consist.**	**Ideal-Type Agencies**
1	0	0	1	0	1	1	.958	.911	Treasury
2	0	0	0	1	1	1	.950	.941	HUD
3	1	0	1	1	1	1	.884	.814	DOD
4	0	1	0	0	1	1	.856	.667	DOL
5	1	1	1	1	0	1	.740	.596	USAID
6	1	1	0	0	0	2	.660	.419	DOC, DOJ
7	0	1	1	1	0	2	.560	.279	DHS, SBA
8	0	1	1	0	0	2	.525	.165	NASA, DOT
9	0	0	0	0	?	-			
10	1	0	0	0	?	-			
11	1	0	1	0	?	-			
12	1	1	1	0	?	-			
13	1	0	0	1	?	-			
14	0	1	0	1	?	-			
15	1	1	0	1	?	-			
16	0	0	1	1	?	-			

Both truth tables indicate a high level of variance in the presence of system action situation and organizational action situation conditions in agencies with positive institutionalization outcomes. For models 1A and 1B, raw consistency values for sufficient sub-set relations range from 1.000 to .662, and PRI consistency values range from 1.000 to .258. Based on the thresholds for raw and PRI consistency already defined based on FSQCA methodological standards, this results in nine minterns associated with 10 agencies coded as having a positive outcome of performance measure adoption and one associated with two agencies with a negative outcome. For models 2A an 2B, raw consistency values for range from .958 to .525, and PRI consistency scores were from .911 to .165. This results in four minterns associated with four agencies with a positive outcome for use of performance information and four associated with seven agencies having a negative outcome. I ran Boolean minimization procedures on each of these truth tables for both

positive and negative institutionalization outcomes for PM and PI. Tables 4.18 and 4.19 present the overall results for models 1A and 1B, respectively, using the truth table at table 4.16. Table 4.20 show the solution for model 2A using the truth table at table 4.17. Minimization of the truth table at table 4.17 for model 2B resulted in a single solution which did not meet the threshold for consistency for being a necessary or sufficient subset for the absence of PI use; therefore, using model 2B, the study did not identify a sufficient causal pathway for agency non-use of performance information.[81]

Table 4.18. Configurations for Agency Adoption of Performance Measures (Model 1A)

	Solution			
	1	**2**	**3**	**4**
Politics and Resource Dependence				
Present		●	●	
Absent	⊖			
Accountable and Empowered Culture				
Present		•		•
Absent				
Capacity for Evaluation				
Present			●	●
Absent				
Mature Performance Measurement System				
Present				
Absent	⊖			
Consistency	.929	.869	.972	.950
Raw Coverage	.323	.408	.470	.434
Unique Coverage	.131	.121	.170	.137
N Agencies	3	3	3	3

Notes:

[81] In QCA methodology, a model that predicts a positive outcome may not be able to predict the absence of the same outcome due to the existence of causal asymmetries.

For the union of all solutions for model 1A, coverage was .913 and consistency was .906, indicating the model is likely very well-specified and explains the outcome well. Ten out of 12 agencies were represented in the four solution sets.

● Core causal condition (present)

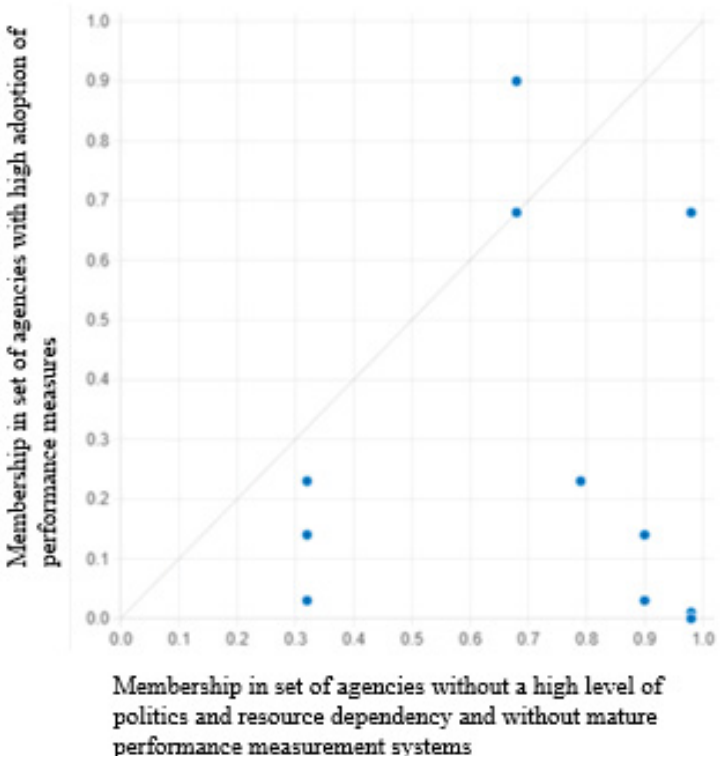
Θ Core causal condition (absent)

● Contributing causal condition (present)

Θ Contributing causal condition (absent)

For model 1A, there are four causal configurations, or solutions, which lead to a positive outcome of adoption of a variety of performance measures for different clusters of federal agencies. Solution 1, exemplified by the HUD, NASA, and Treasury cluster, is characterized by the perceived absence of politics and resource dependency (~P&R) and the perceived absence of a mature performance measurement system (~M). The set relationship between this configuration and the set of agencies adopting a variety of performance measures is consistent (.929) with being one of necessity, as clearly seen in the scatterplot in figure 4.18, in which the great majority of membership scores for the configuration of conditions (X axis) are greater than the associated membership score for the PM outcomes (Y axis).

Figure 4.18. Model 1A Solution 1: XY Plot – Condition ~PR*~M, Outcome PM



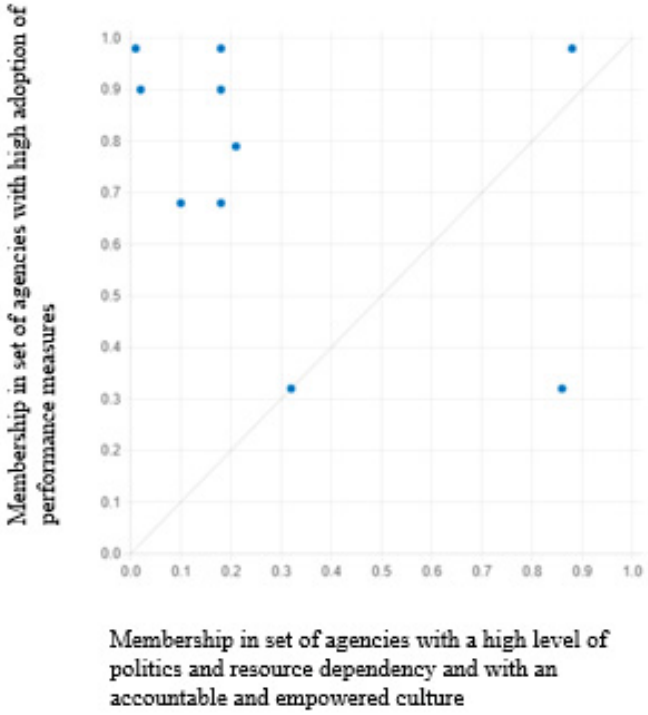
Consistency	.929
Raw Coverage	.323
Unique Coverage	.131
N Agencies	3

One interpretation of this result is that agencies which are less exposed to the influence of politics and resource dependency and lack a mature system to measure their performance will adopt a variety of performance measures, even without the presence of an accountable and empowered culture or the capacity for evaluation. This makes sense because for some agencies, the absence of political pressure may encourage (or at least not discourage) agency leaders to develop performance measures for internal performance reasons, vice compliance. That is not to claim this would be true for all agencies. For agencies in this cluster, the absence of performance measures, or the presence of a few measures, would seem to be easily recognized by senior leaders and used as justification to develop more and/or more sophisticated performance measures.

An analysis of set relations for this solution points to the set of agencies adopting a variety of performance measures being a subset of the set of agencies without a high level of P&R and without advanced performance measurement systems, making the combination of these two conditions necessary for a positive institutionalization outcome, all else being equal. Along with the relatively low coverage parameter (.323 raw and .131 unique coverage of the outcome), this suggests there are almost certainly other conditions at play in these three agencies which, when active in the context of this necessary configuration, may help to explain these agencies' adoption of a variety of performance measures. In other words, the absence of political and resource dependency and the absence of mature measurement systems may in some respect be "necessary," but the necessity is rather trivial due to low coverage, as guidance from Schneider and Wagemann (2013, 144-146) would suggest. To summarize findings for solution 1, the set relations for this solution indicates PSMs in these agencies may have substantial intrinsic motivation to measure agency performance, even without being directed to do so, recognize situations when developing new measures would add value to performance, and feel relatively less constrained by political pressures that might otherwise hold them back from institutionalizing new measures.

Solution 2, which includes DOC, DOJ, and USAID, is characterized by the presence of P&R and an accountable and empowered culture (C). The set relationship between this configuration and the set of agencies adopting a variety of performance measures is consistent with sufficiency, as clearly seen in the scatterplot in figure 4.19, in which the great majority of membership scores for the configuration of conditions (X axis values) are less than the associated membership score for the PM outcomes (Y axis values).

Figure 4.19. Model 1A Solution 2: XY Plot – Condition P&R*C, Outcome PM



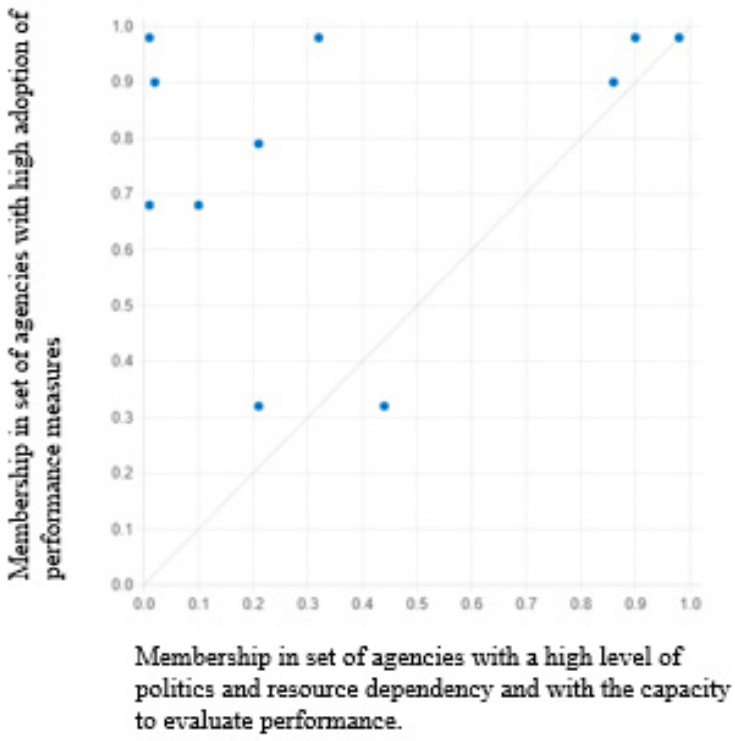
Consistency	.869
Raw Coverage	.408
Unique Coverage	.121
N Agencies	3

These results suggest DOC, DOJ, and USAID have adopted a variety of performance measures in response to external political and resource pressures. The evidence also indicates the presence of accountable and empowered cultures in these agencies has enabled them to respond positively to external political and resource pressures and led agency performance system managers to institutionalize diverse measures of performance, likely those of importance to political leaders and Congress more than operational managers within the agency. That capacity for evaluation (X) and M were not conditional elements of this causal configuration further suggests these agencies evaluate measures of performance primarily to respond to the needs of external stakeholders, but perhaps not to support the informational needs of line managers. Together,

P&R and C have been a sufficient combination of conditions for DOC, DOJ, and USAID to adopt a variety of performance measures.

Solution 3, which describes DOD, HHS, and USAID, is similar to solution 2, except it replaces X for C, and also identifies P&R and X as core conditions. The set relationship between the intersection of these sets and the set of agencies adopting a variety of performance measures is clearly consistent (.972) with one of sufficiency, as shown in the scatterplot in figure 4.20, in which the great majority of membership scores for the configuration of conditions (X axis) are less than the associated membership score for the PM outcomes (Y axis).

Figure 4.20. Model 1A Solution 3: XY Plot – Condition P&R*X, Outcome PM



Consistency	.972
Raw Coverage	.470
Unique Coverage	.170
N Agencies	3

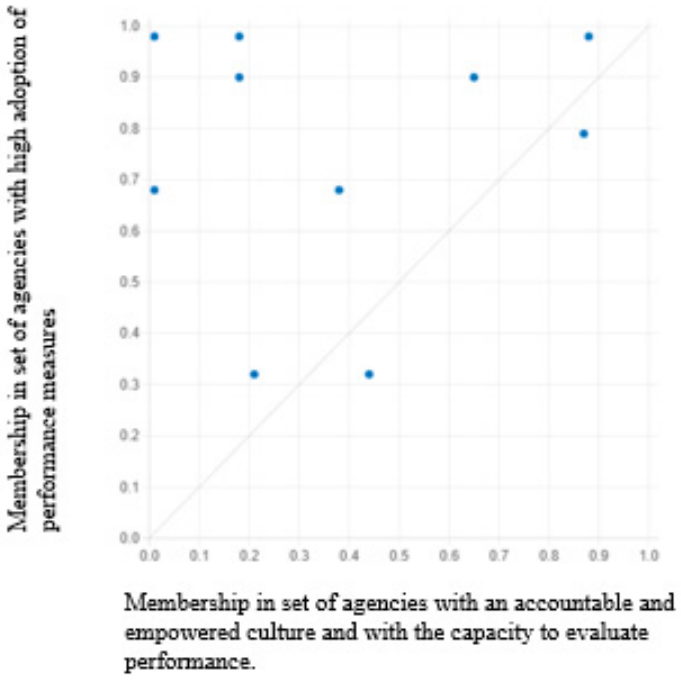
USAID makes a repeat appearance in this configuration. When analyzing fuzzy set relationships, a case may appear in more than one configuration based on its characteristics. The common condition of P&R in both configurations suggests it may play an outsized role for performance management at USAID. This is not at all surprising, given USAID's role as an administrator of over $19 billion in foreign aid[82] and well-publicized recurring political machinations in Congress to reduce foreign aid budgets. DOD would also seem to be an exemplar for external political influence on its performance management system, because of enduring Congressional interest in defense budgets to channel money to congressional districts or Presidential interest in using DOD resources for non-defense purposes, such as building a border wall. HHS has arguably also been the focal point for disproportionate congressional interest in the budgets of social programs, the core of the department's mission identity, as political control of Congress and Administrations shifts back and forth between political parties with different social agendas. The configuration suggests USAID, DOD and HHS have made previous substantial investments into capacity for evaluating performance. Together, P&R and X have been sufficient for these agencies to institutionalize the adoption of a variety of performance measures. With enough investment in capacity and top-down political pressure, a culture of accountability and empowerment and mature performance measurement systems may be less of a requirement for adopting a variety of performance measures.

Solution 4 describes the sufficient causal configuration that led DHS, SBA, and USAID (again!) to adopt a variety of performance measures in their performance management programs. The set relationship between this configuration and the set of agencies adopting a variety of

[82] https://www.usaid.gov/news-information/fact-sheets/fiscal-year-2020-development-and-humanitarian-assistance-budget-request

performance measures is one of sufficiency, as shown in the scatterplot in figure 4.21, in which the great majority of membership scores for the configuration of conditions (X axis) are less than the associated membership score for the PM outcomes (Y axis).

Figure 4.21. Model 1A Solution 4: XY Plot – Condition C*X, Outcome PM



Consistency	.950
Raw Coverage	.434
Unique Coverage	.137
N Agencies	3

The results suggest previous management focus on improving the results orientation of agency culture and investments in evaluation capacity best explain why DHS, SBA, and USAID have adopted a wide range of performance measures in their performance management programs. DHS and SBA are uniquely aligned to this cluster. USAID's presence in this cluster is explained by the fact that it also appears in cluster 2 and 3, in which C and X were contributing and core conditions, respectively, in sufficient configurations.

In model 1B, which models the causal pathway leading to non-adoption of performance measures, I negate the outcome of adoption of performance measures by subtracting all PM fuzzy set data from 1 and, as with model 1A, I assume based on theoretical expectations that P&R may be present or absent when the performance measures are not adopted. For the other conditions in the model, I assume based on theory and the literature review that their absence (vice presence as in model 1A) will contribute to the outcome of non-adoption. Simplifying the truth table at table 4.16 using Boolean minimization, a single minimally-consistent (.804) sufficient configuration emerges to explain the set of conditions which together may explain the non-adoption of a variety of performance measures by federal agencies. The results are at table 4.19. The relatively high coverage parameters (.596), equal because there is only one configuration in both the parsimonious and intermediate solutions generated by FSQCA, provide some assurance that the configuration provides a substantial part of the explanation for the outcome of non-adoption, at least in the two agencies providing evidence, which were DOL and DOT, and to the degree that the configuration is indeed consistent with a subset relationship of sufficiency.

Table 4.19. Configurations for Agency Non-Adoption of Performance Measures (Model 1B)

	Outcome
	1
Politics and Resource Dependence	⊖
Accountable and Empowered Culture	
Capacity for Evaluation	⊖
Mature Performance Measurement System	●
Consistency	.804
Raw Coverage	.596
Unique Coverage	.596

N Agencies	2

Notes:

Notes:

For this solution, coverage of the outcome was .596 and consistency was .804, indicating the model met minimum standards for consistency with being a sufficient causal configuration, and partially explains the outcome of agency non-adoption of performance measures. The coverage value shows the partial membership of other agencies in this configuration.

● Core causal condition (present)

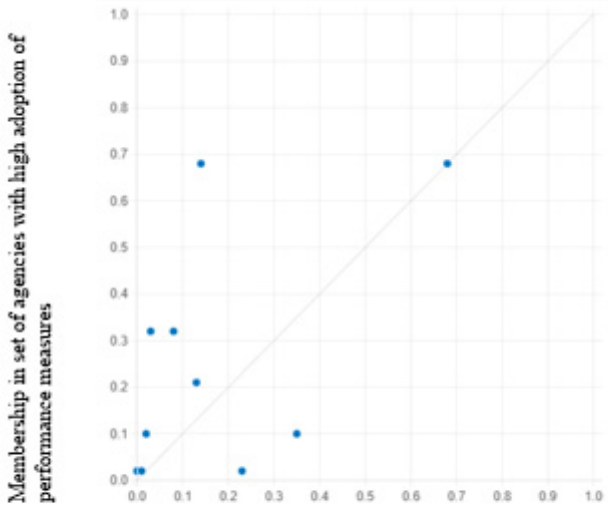
Θ Core causal condition (absent)

● Contributing causal condition (present)

Θ Contributing causal condition (absent)

The subset relationship between this configuration and the set of agencies failing to adopt a variety of performance measures is barely consistent (.804) with being one of sufficiency, as seen in the scatterplot in figure 4.22, in which the majority of membership scores for the configuration of conditions (X axis) are less than the associated membership score for the PM outcomes (Y axis). The generally low X values are due to the negation of P&R and X, as one would expect.

Figure 4.22. Model 1B Solution 1: XY Plot – Condition ~PR*~X*M, Outcome ~PM



Consistency	.804
Raw Coverage	.596
Unique Coverage	.596
N Agencies	2

This finding is interesting for two reasons. First, it seems to empirically confirm findings in previous positive outcome configurations that the presence of a mature performance measurement system acts as a negative factor on agency adoption of additional performance measures, and its absence, as in the case of solution 1 of model 1A, may encourage it. This finding contradicts scattered empirical research as noted in chapter 2, which has found the opposite to be true. Because M is an autoregressive variable in the quantitative model, it exerts a similar "carry over" influence in the qualitative model. Therefore, the absence of politics and resource dependency, combined with the absence of capacity for evaluation, in agencies with mature performance measurement systems, may better explain why those agencies fail to adopt a wide range of performance measures.

With routinely used performance measures, little political pressure, and inadequate resources for performance evaluation, it should not surprise anyone that federal managers would feel little incentive to expand the number and type of indicators of performance to measure. Why make more work for your organization when there is little demand signal from external stakeholders, few resources to expand activity, and current measurement systems seem to be adequately mature? GPRAMA institutionalization in terms of adoption of performance measures is therefore hindered by the combination of the absence of politics and resource dependency (in the case of some agencies), the absence of evaluation capacity, and the presence of a mature performance measurement system. Whether this is a fair characterization of DOL and DOT performance system manager attitudes will be considered in the next section.

For model 2A, the Boolean minimization revealed two causal configurations that were consistent with a sufficient subset relationship to explain the use of performance information (PI) by managers in federal agencies. Based on theory and the literature review, model 2A assumes P&R could be either present or absent, and that C, credible leadership commitment (L) and X are present when there is a positive PI use outcome.

Table 4.20. Configurations for Agency Use of Performance Information (Model 2A)

	Solutions	
	1	**2**
Politics and Resource Dependence	⊖	
Accountable and Empowered Culture	⊖	⊖
Capacity for Evaluation		●
Credible Leadership Commitment	●	●
Consistency	.958	.893
Raw Coverage	.152	.182
Unique Coverage	.090	.142
N Agencies	1	1

Notes:

● Core causal condition (present)

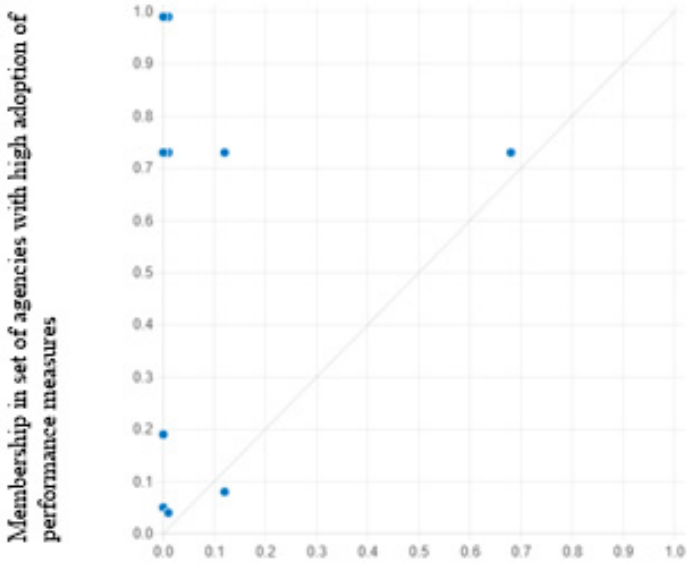
Θ Core causal condition (absent)

● Contributing causal condition (present)

Θ Contributing causal condition (absent)

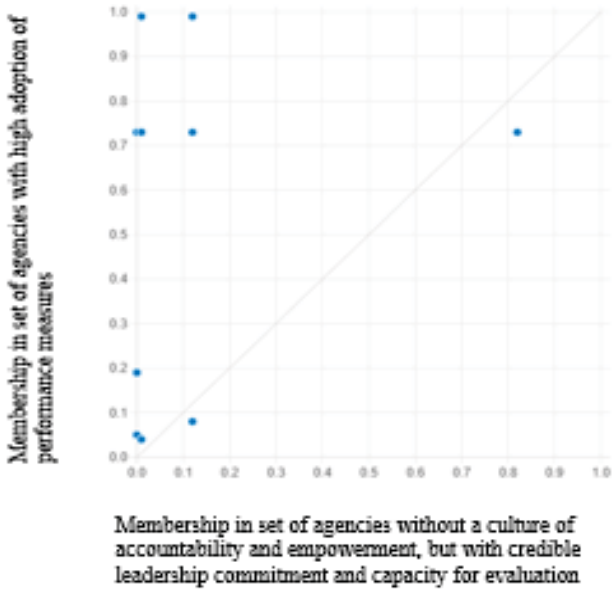
The Treasury department is the ideal-type agency for the solution 1 configuration, and DOD is the ideal type for solution 2. In each case, the evidence from the survey of federal performance system managers points to their perceptions that their agency lacks an accountable and empowered culture, but is able to make use of performance information based on the presence of credible leadership commitment (both Treasury and DOD), the lack of political pressure (Treasury), and the presence of robust evaluation capacity (DOD). The set relationship between the configurations in solutions 1 and 2, and the set of agencies that make extensive use of PI, is strongly consistent with sufficiency, as seen in the scatterplots in figures 4.23 and 4.24, in which the great majority of membership scores for the configuration of conditions (X axis values) are less than the associated membership scores for the PM outcomes (Y axis values).

Figure 4.23. Model 2A Solution 1: XY Plot – Condition ~PR*~C*L, Outcome PI



Consistency	.958
Raw Coverage	.152
Unique Coverage	.090
N Agencies	1

Figure 4.24. Model 2A Solution 2: XY Plot – Condition ~C*L*X, Outcome PI



Consistency	.893
Raw Coverage	.182
Unique Coverage	.141
N Agencies	1

The low raw and even lower unique coverage values for both solutions, however, suggest each solution explains only a very small part of the positive outcome of PI use. Together with three other generated configurations which by themselves did not meet the defined consistency thresholds but included DOL and HUD, the two combined sufficient solutions achieve an outcome coverage value of only .507. This means DOL and HUD share some of these same conditions with Treasury and DOD to a degree, but also that there must be other conditions that affect their PI use that are not part of model 2A. Therefore, the key findings regarding causal configurations leading to agency use for performance information are, contrary to the results of the quantitative research, that when considered as an element of a causal configuration, a culture of accountability and empowerment is neither necessary nor sufficient to result in PI use by federal agencies, and that when agencies lack a culture of accountability and empowerment, the

presence of credible leadership commitment and investments in evaluation capacity may help to foster the use of performance information.

Given how challenging it is for federal leaders to change organizational cultures, this finding offers some encouragement that the means to foster use of performance information are available to them short of having to revolutionize agency cultures, which can take decades to implement and sediment. However, because the evidence base in the analysis of model 2A is thin, much more research is needed to validate this claim and to identify other possible conditions that systematically lead agencies to use PI to help managers make decisions. Finally, as noted, the QCA results did not yield any causal configurations that could consistently explain the outcome of non-use of performance information. This means a much more inclusive and comprehensive model of causality must be proposed to explain agency non-use of performance information, likely using a much larger dataset given the wide range of possible conditions that could combine to lead to negative outcomes. Table 4.21 summarizes the results of qualitative sub-study 1.

Table 4.21. Summary of Causal Configurations for GPRAMA Institutionalization Outcomes

Model	Outcome	Solution	Configuration	Ideal-Type Federal Agencies
1A	PM Adoption	1	~P&R*~M	HUD[1], NASA[1], Treasury[1]
		2	P&R*C	DOC, DOJ, USAID
		3	P&R*X	DOD, HHS, USAID
		4	C*X	DHS[1], SBA[1], USAID
1B	PM Non-Adoption	1	~P&R*~X*M	DOL, DOT
2A	PI Use	1	~P&R*~C*L	Treasury[1]
		2	~C*L*X	DOD
2B	PI Non-Use	None	None	Inconsistent membership patterns

~ Symbol for "absence of"
* Symbol for Boolean "and"
[1] Indicates one or more PSMs in the agency also participated in an interview.

Qualitative Sub-Study 2: Case Studies of Five Federal Agencies

For the second qualitative sub-study, interviews with six federal PSMs provided insights on organizational contexts that help to explain how system action situation and organizational action situation conditions influenced the relative level of institutionalization of GPRAMA in their agencies. PSMs interviewed worked for DHS, HUD, NASA, SBA, and Treasury. The 2021 survey responses were used to help elicit responses to some questions. Drawing from table 4.13, table 4.22 focuses on the descriptive results of the five agencies from the 2021 PSM survey. Values are the mean of individual PSM index scores for the number of different types of institutionalization outcomes observed as present, and the number of indicators for system and organizational conditions observed as present to a "great" or "very great" extent in their agencies.

Table 4.22. Selected Agency Scores from 2021 Federal Performance System Manager Survey

	Institutionalization Outcomes		**System and Organizational Conditions**				
	PM	**PI**	**P&R**	**C**	**L**	**X**	**M**
DHS	4.40	4.40	2.80	4.60	4.40	4.40	7.00
HUD	6.00	8.00	3.00	2.00	1.00	4.00	6.00
NASA	4.00	4.50	2.50	6.50	6.00	3.50	5.50
SBA	5.00	6.00	2.00	4.00	7.00	4.00	8.00
Treasury	4.00	6.00	3.00	3.00	5.00	1.00	5.00
MAXIMUM	6.00	8.00	6.00	6.50	7.00	5.00	8.00
MINIMUM	3.00	4.40	2.00	2.00	1.00	1.00	5.00
MEAN	4.62	5.87	3.65	4.09	4.62	3.50	6.55

Notes:
PM Adoption of Performance Measure
PI Use of Performance Information
P&R Politics and Resource Dependency
C Accountable and Empowered Culture
L Credible Leadership Commitment
X Capacity for Evaluation of Performance
M Measurement System Maturity
Maximum, Minimum and Mean refer to all agencies surveyed, not just those in the table.

Table 4.21 showed HUD, NASA, Treasury, DHS, and SBA as ideal-type agencies for two different causal configurations that led to a positive outcome for adoption of performance measures. At HUD, NASA, and Treasury, the QCA findings for model 1A showed the absence of P&R and the absence of M, combined, were a *necessary* causal configuration for the adoption of performance measures, but there was not enough evidence in the survey data to identify sufficient causal configurations for these agencies.

This result begs the questions of why these conditions in the necessary configuration were generally absent vice present, and what other present conditions or intra-agency dynamics were able to take advantage of the relative absence of these two conditions to bring about a positive institutionalization outcome in these two agencies? At Treasury, model 2A showed the absence of P&R, the absence of C, and the presence of L were a sufficient causal combination to achieve a positive outcome for the use of performance information. But, what characteristics of senior leadership at Treasury make it so credible to agency managers and significant for encouraging them to use performance information, particularly when the 2021 survey index for C (3.0) was well below the federal average (4.09)? At DHS and SBA, QCA findings for model 1A showed the combination of C and X was sufficient for a positive outcome for performance measure adoption, which is in line with the results of both the quantitative analysis and mainstream theory and evidence offered in the literature. But, what specific features of their organizational cultures and their capacity to evaluate performance affected their institutionalization of a variety of performance measures, and how did they do so? Concerning both roads to performance measurement adoption, did the interviews provide evidence of possible alternative explanations for these four agencies? These questions helped to focus the

analysis on the evidence generated during the interviews with the PSMs at the five federal agencies.

Sustainable institutionalization of reform policies must have causes, but in some cases, these causes can be obscure and difficult to characterize. Analysis of the condition and outcome subset relationships for the institutionalization of a wide range of performance measures in sub-study 1 identified a definite but perhaps trivial relationship of necessity for the combination of the absence of P&R and M and a positive outcome for performance measure adoption at HUD, NASA, and Treasury. This causal configuration seems to support some theoretical predictions, as noted in chapter 2, that the absence of politics and resource dependency are favorable conditions for some public organizations to adopt performance measures.

Necessary configurations, however, only represent a backdrop for the interplay of conditions which are present but perhaps difficult to detect using a survey instrument. The survey data alone did not provide consistent evidence of any combination of present conditions to explain sufficient causal relationships for these agencies. Interviews with PSMs at all three agencies generated supplemental evidence not captured through the survey to characterize the necessary conditions and explain the influence of other conditions in those organizations which might be sufficient in some combination to explain positive outcomes for performance measure adoption and, in the case of Treasury, the use of performance information. Clustered summary tables (Miles and Huberman 1994, 182) at appendices Q, R, and S present the case evidence from these interviews with PSMs at HUD, NASA, and Treasury respectively.

Analysis of the condition and outcome subset relationships in the QCA identified a combination of the presence of C and X was consistently sufficient to bring about a positive outcome for the adoption of performance measures at DHS and SBA. This causal configuration

seems to support some theoretical predictions, as noted in chapter 2, that the presence of an accountable and empowered culture and the presence of capacity for performance evaluation are favorable conditions for agencies to adopt a wide range of performance measures. Clustered summary tables at appendices T and U (Miles and Huberman 1994, 182) summarize case evidence from DHS and SBA.

In addition, the interviews included questions to gauge PSM attitudes, intrinsic motivation, and biases concerning performance management. The inferential model of this study treats these as conditions within the individual action situation that may mediate the effects of organizational action situation conditions and incentive structures on manager decisions to adopt performance measures and use performance information. Because the number of PSMs available to interview was very small and therefore not sufficient to reliably model interaction effects of manager attitudes and biases with organizational conditions on GPRAMA institutionalization outcomes at the agency level, the results are provided in these cases simply to facilitate the interpretation and evaluation of the survey and interview results on the other topics because PSM attitudes and biases could affect responses to all of the other survey and interview questions. The following five case study narratives compare and evaluate the evidence from the 2021 survey, the QCA, and the PSM interviews for the five agencies which participated in the interviews.

Case Study #1: Department of Housing and Urban Development (HUD)

Survey and QCA Findings. At HUD, the 2021 survey index scores for PSM perspectives on PM (6.00) and PI (8.00) from table 4.22 were both at the maximum level for all federal agencies studied. The index score for P&R (3.00) was below the federal mean of agencies surveyed (3.65), as was M (6.00/6.55). The QCA results conform with the survey data. Because

the absence of P&R and M was shown in the QCA to be part of the necessary causal configuration for HUD, this suggests that relative agency autonomy from political interference and the relative under-development of a performance measurement system facilitates the adoption of performance measures. The survey data on the system and organizational conditions at HUD show condition index scores consistently below the federal average for all conditions except for capacity to evaluate performance. Therefore, from the survey data alone, it is difficult to identify sufficient causal configurations that explain HUD's very high institutionalization outcome indices. The interviews provided useful supplemental evidence.

Interview Findings. According to the interviews with HUD PSMs, the HUD PMS enjoys considerable protection from political and resource dependence concerns. Political appointees at HUD normally serve for less than two years. As a result, senior career officials often feel they can "wait them out" if they propose significant changes. In addition, the PSM interviews indicated HUD's most senior leaders closely guard their prerogatives to decide on agency performance measures tied to the strategic plan, as well as the plan itself, and have demonstrated their ability to adeptly respond to a wide range of external partners and stakeholders, to include industry groups and public housing associations. Whether the leaders are adept with their stakeholder engagements, or are simply unresponsive to external stakeholders as noted by GAO as recently as 2016 (U.S. Government Accountability Office 2016, 17-18), the political top cover provided by HUD's most senior leaders provides another degree of insulation for HUD's PMS system from political influences.

The relatively low index score for M in the HUD PSM survey may be explained by the interview findings that measurement standardization is limited to within program offices and enterprise-wide standardization is hindered by poor IT interoperability, and by the repeatedly

documented difficulty that state and local mission partners have when trying to apply HUD's performance measurement frameworks at the local level. This evidence conforms with the literature on implementation of performance management in an environment of fiscal federalism, in which federal programs are not administered directly or via hierarchical systems, but instead via third party networks (e.g. Frederickson and Frederickson 2006). It seems doubtful either condition would allow HUD managers to obtain program level performance information that is valid or timely for decision making or which satisfy GPRAMA reporting requirements. While there is some PMS integration between strategic planning, budgeting, and mission activities for agency priority goals (APGs), it is still lacking for non-priority goals, a much larger goal set. This suggests that management is prioritizing the use of limited staff resources in the face of high transaction costs for measuring non-APGs. Based on this triangulation of evidence, the HUD interviews support the QCA model 1 solution 1 results.

Based on evidence from the HUD PSM interviews at appendix Q, system action situation and organizational action situation conditions at HUD that the literature predicts should foster the adoption of performance measures are: 1) the advancement of capacity for evaluation at the program office level and with the support of the Office of Policy and Research and Division of Strategic Planning and Performance Development, 2) the legacy of the HUDStat experience in developing and using performance measures that track program outcomes, and 3) the innovative application of risk management frameworks in the strategic planning and performance evaluation process.[83]

[83] The interviews revealed that under the Trump Administration, HUDStat became a legacy efficiency project, but its effects had become somewhat sedimented in HUD performance measurement practices. Previously, HUDStat had been highlighted as ". . . a series of monthly, data-driven meetings led by the Secretary. . . " and an important part of HUD's transformation initiative that was focused on measuring HUD progress against their high priority performance

PSM Attitudes and Biases. HUD PSMs have definite views about the perspectives of line managers at HUD and how they perceive GPRAMA, and HUD processes to implement it. Without a legal mandate, PSMs believed most managers at HUD would not volunteer to participate in a performance reporting process. They believe HUD managers see it as a compliance exercise, and believe that the actual performance reports, once published, are not read due to lack of public interest or publicity. Nevertheless, the PSMs saw instrumental organizational benefits for HUD from GPRAMA, especially the effect of encouraging HUD leadership to create quarterly meetings that bring together all program offices to discuss their performance, thus reinforcing the "One HUD" theme, an important part of the department's strategic communications program. The PSMs felt that, without GPRAMA, new program office "fiefdoms" would emerge within HUD. They indicated that GPRAMA requirements likely do help HUD to improve performance, but stated it was difficult to know whether or not it saved HUD any money, particularly because almost all of HUD's internal budget is for manpower.

Case Study #2: National Aeronautics and Space Administration (NASA)

Survey and QCA Findings. At NASA, the 2021 survey index scores in table 4.22 for PM (4.00) and PI (4.50) were slightly below the mean for all federal agencies (4.62 and 5.87, respectively), but the index values for P&R (2.50) and M (5.50) were even lower than the federal averages (3.65 and 6.55, respectively) in relative terms, which helps explain the membership of NASA in this causal configuration. In the QCA results, NASA, along with HUD and Treasury,

goals (U.S. Government Accountability Office 2011b, 16). The Obama Administration archives associated HUDStat with a joint HUD-VA program that reduced veteran homelessness by 47 percent from 2010-2016 (https://obamaadministration.archives.performance.gov/content/veteran-homelessness-reduced-47-over-6-years.html).

was an ideal-type agency for a consistently necessary causal configuration including the absence of both P&R and M. The interpretation of this configuration for HUD applies equally to NASA.

Interview Findings. The interview findings mostly conform with the survey and QCA results. Reviewing the major findings from the interviews with NASA PSMs at appendix R, it is clear many aspects of NASA's response to GPRAMA have been strongly influenced by its self-identification as a "research and development" (R&D) agency. PSM and line manager preferences for performance measurement are strong but fundamentally different from the outward-facing, transparent, and accountability-based performance management framework of GPRAMA. As an R&D agency, managers would prefer to make routine use of performance information about their projects and programs, even in the absence of GPRAMA (unlike at HUD). But the performance data would only be for internal consumption by program and project managers to improve tactical decision making. NASA has gone beyond this level of performance reporting in its institutionalization.

Because GPRAMA is the law, NASA has adapted and accommodated the OMB-issued guidance to ensure it stays in compliance, and the PMS organization and NASA senior leaders have made a good faith effort to make use of existing performance measurement processes and resources at the program office level to satisfy GPRAMA reporting requirements. The interview evidence points to a trend towards deeper institutionalization of GPRAMA-style measures over time, especially outcome measures. While there are certainly external political and resource influences on NASA programs, NASA PSMs believe GPRAMA institutionalization is not much of a concern for anyone outside NASA, except for OMB. NASA PSMs believe Congress shows little or no interest in NASA performance reporting under GPRAMA. The ex-Trump Administration requirement to exclude all strategic planning goals related to the climate seems to

be the only overt effort to interfere with NASA's strategic planning and performance management processes, and there is no evidence that any resource or disciplinary consequences were associated with it.

Unlike HUD, where all critical decisions about performance management are taken at the very highest levels of the department, NASA's approach is much more decentralized, and empowers its bureaus, mission directorates and support offices to take the lead for developing measures, evaluating their own performance, and sending performance reports to NASA headquarters for consolidation, but not review or revision. Along with the only very recent addition of targets to the 2020 performance plan, this decentralized approach to managing performance measurement probably explains the relatively low PSM assessment on the 2021 survey about the maturity of its performance measurement system due to its lack of standardization at the enterprise level and lack of experience with target measures. Nevertheless, NASA has effectively institutionalized outcome measures, which the literature proposes are the most difficult for any agency to operationalize in a PMS. And, there have been recent improvements in PMS integration at the headquarters level: NASA leadership holds monthly Baseline Performance Reviews which include both GPRAMA and non-GPRAMA related performance updates. NASA PSMs believe these monthly meetings have substantially raised the credibility of senior leadership commitment to performance management and have created a strong incentive for the subordinate units to participate in the meetings.

Evaluating the evidence on NASA institutionalization of GPRAMA from the interviews with the NASA PSMs, the most likely conditions that, in some combination, likely lead to positive outcomes for a high level of performance measure adoption are: 1) an organizational culture that is naturally focused on R&D project and program performance and staffed with

highly-educated professional scientists and engineers who are likely comfortable with a high degree of quantification in business practices,[84] 2) credible leadership commitment developed by assertive agency leadership to rally program offices in monthly BPR meetings, take program reports seriously by personal engagement in meetings, and hold program managers accountable for their performance, and 3) despite the relatively low index value for M in the 2021 survey, the reality that NASA has made strides in operationalizing outcome based performance measures and recently started to apply standard techniques of target setting to its performance goals. The interview results paint a more favorable picture for the maturity of the NASA performance measurement system than the NASA PSM responses to the 2021 survey did.

PSM Attitudes and Biases. In evaluating the value of GPRAMA for NASA, PSMs at NASA, like those at HUD, felt its real value was exclusively instrumental. One respondent pointedly indicated that "The useful part is doing the dance. The end products aren't particularly useful." By serving as a forcing function for self-evaluation, GPRAMA brings mission directorates and headquarters together in one place on a regular basis, which permits sharing of information, feedback, and occasionally much-needed "uncomfortable conversations." The PSMs felt GPRAMA reporting may help to improve performance, but probably not to reduce costs. PSMs believed institutionalizing GPRAMA reporting in NASA program offices was not easy. The NASA PSMs, who are strong advocates of performance management generally, half-jokingly concluded with the comment that "it's taken a long time to beat them into submission."

Case Study #3: Department of the Treasury

[84] Additional evidence for this claim comes from the 2021 survey results, in which NASA's index for accountable and empowered culture was at the highest level of the range for the whole federal government.

Survey and QCA Findings. At the Treasury Department, the 2021 survey of PSMs showed that the performance measures adoption index (PM) (4.00) was slightly below the federal mean (4.62) but, like NASA, both P&R (3.00) and M (5.00) were well below the federal mean (3.65 and 6.55 respectively), in line with the QCA results. As with NASA and HUD, the causal configuration for PM describes a consistently necessary subset relationship at Treasury, so an effort must be made to inductively postulate sufficient causes for performance measure adoption at Treasury based on the evidence from the interviews.

The 2021 PSM survey also showed that Treasury's PI index score (6.00) was above the federal mean (5.87), P&R (3.00) and C (3.00) were below the federal mean (3.65 and 4.09, respectively) and L (5.00) was above the federal mean (4.62), also in line with the QCA results. This is potentially an important finding, which suggests that strong and credible senior federal leaders who are acknowledged by managers as committed to performance management can have an outsized impact on their agencies' use of performance information, *even when the agency's culture is not especially accountable or empowered*, as long as there is an absence of political interference or resource dependency for the agency's PMS. The evidence from interviews of PSMs at the Treasury Department in appendix S aligns with the evidence from the surveys and supports this finding.

Interview Findings. Whereas HUD PSMs invoked the grant-facilitating nature of their mission, and NASA PSMs invoked the unique research and development focus of the organization, Treasury PSMs claimed no unique mission-based exceptionalism for Treasury's PMS. This makes the case of Treasury somewhat more generalizable to the majority of other federal agencies who likewise provide direct services to the government and citizens. The structure of Treasury's PMS appears to strike a balance between the very top-down,

headquarters-managed hierarchical structure in HUD and the very decentralized structure in NASA, in which the headquarters serves a ministerial and administrative role, but defers critical decisions to bureaus. The structure of participants managing the PMS is very similar to the other agencies, with a COO (the Deputy Secretary), a PIO (the Assistant Secretary for Management) and a Deputy PIO (Director of Office of Strategic Planning and Performance Improvement), goal leads, and several dozen performance leads in the bureaus and offices, but also includes conceptually the chief risk officer and the chief data officer. Strategic planning, budgeting, and strategic management routines are integrated business management processes across Treasury, and responsibility for governance decisions is shared between headquarters and bureaus and offices. There is a high level of voluntary participation by senior executives from all bureaus in Treasury's Quarterly Performance Review (QPR) process, with PSMs estimating that as many as 140 senior executives participate in an annual cycle of four performance review meetings, and a broad base of leadership participation is a sign of deep institutionalization.

As with HUD and NASA PSMs, PSMs at Treasury reported hardly any political influence on, or resource dependence by, the Treasury PMS. When asked about the influence of stakeholders who shaped the development of the Treasury PMS, the only external entity mentioned was OMB, which PSMs believed "gave Treasury a lot of latitude." Key stakeholders identified were the Assistant Secretary for Management, the DPIO and staff, and especially the Deputy Secretary, who perceives GPRAMA as a useful management tool. PSMs did not rule out the possibility of hidden political influences, but could not recall specific instances they had observed or heard about. The PSMs noted there were always competing interests on specific departmental policies and programs by OMB and Congress, but these disputes are seen as legitimate parts of the policy process and not an effort to influence performance reporting. The

preponderance of evidence aligns Treasury with HUD and NASA in the set of agencies which enjoy minimal political interference in, or risk of budgetary manipulation of, their PMS. This gives leaders and PSMs a free hand to develop PMS systems that are driven mostly by internal information needs.

Agency culture, or rather its relative absence, was identified in the QCA as a condition in the sufficient configuration leading to agency manager use of PI. This is not meant to imply that the absence of C contributes to the outcome per se, but that its relative absence was a simple fact in the case of Treasury and that other conditions in the configuration overcame the adverse influence of the absence of C to bring about the positive outcome. PSMs believe that managers in Treasury do collaborate regularly and are held accountable for achieving results. They cited some examples of programs which had improved over time as evidence of a results-focused culture, such as the reduction in the number of paper checks in the early days of GPRAMA and ongoing efforts to improve efficiency by collaborating with the IRS to reduce the overall number of checks sent to beneficiaries via electronic deposit. The 2021 survey results painted a mediocre picture, however, with Treasury's culture index of performance culture indicators considerably below the federal average. Evidence of improvement in program performance is not the best proof of an agency's cultural climate of accountability and empowerment because many conditions other than culture can bring that about. In this case, the survey data likely present a more reliable picture of Treasury's organizational culture. The relative silence of PSMs on culture-related questions during the interviews spoke volumes.

On the other hand, the high survey index for credible leadership commitment, which was above the federal mean, was well-supported by interview evidence. PSMs indicated senior leaders continuously communicate their expectation for ongoing improvement in practice

development and goal setting. Treasury leaders are concerned about the performance culture in the department and have developed an index for measuring performance culture which enables the department to generate trending data for analysis of actionable improvements. The noteworthy participation of over 140 senior executives in QPRs has already been noted. PSMs were enthusiastic about the quality of the leadership discussions and the sheer number of issues that are addressed at QPRs. They noted that annual results on the Federal Employee Viewpoint Surveys also pointed to workforce perceptions of strong and credible leadership commitment to performance management. The high survey index score for L appears to be well-substantiated and in the case of Treasury is likely the core condition in the sufficient causal configuration resulting in manager use of PI at Treasury.

The absence of M in the case of Treasury was a core condition for agency adoption of performance measures as it was for HUD and NASA, and its survey index score for M was lower than those of both other agencies in this group and well below the federal mean. Thus, we can glean perhaps the best interpretation of why a low M index could lead to a positive institutionalization outcome by examining the condition closely at Treasury. From the 2021 survey results, there were three indicators of maturity absent at Treasury: the timeliness of performance data for decision making, standard definitions to measure organizational performance, and achievable goals. The interview confirmed this data but also pointed to other aspects of PMS maturity such as an institutionalized and standardized approach to developing measures. PSMs noted Treasury was "not too standardized" in how it analyzes performance information, noting "it depends on how the individual analyst approaches it." On the other hand, Treasury instituted an annual process to assess the validity and quality of its performance measures, and the criteria for evaluation are themselves standardized. And, between them and

HUD and NASA, Treasury is the only agency whose PSMs reported the department promulgated an official policy on measuring and reporting organizational performance. Finally, interview results indicated performance coordinators in bureaus had been reporting increasing use of PI by line managers at lower levels over the past 8 years. All of this creates an image of an agency going through a learning and developmental process and gradually maturing its PMS. Its relatively low index score for this condition in the survey may mask important trends happening under the surface that are improving how Treasury measures its performance. An element of self-criticism by Treasury PSMs may also explain what could be an underestimation of their agency's progress in institutionalizing routines to develop worthwhile performance measures. I find the results more likely confirm positive propositions about M in the literature than they invalidate them, especially since in the case of the lowest-scoring agency, the agency average scores for 5 out of 8 survey questions were still answered in the affirmative.

To summarize the findings for Treasury, I accept the necessary causal configuration for adoption of performance measures in the case of Treasury, but acknowledge the QCA and survey results cannot explain which configuration of conditions is *sufficient* to explain it, and also that the maturity of Treasury's performance measurement system is possibly underestimated in the data. The interview data point to several likely conditions that have helped Treasury adopt a wide range of performance measures that are valid and reliable. These include 1) a close linkage between synchronized management processes of strategic planning, annual budget formulation, and ongoing strategic management routines, 2) a history of success with tracking program performance improvements such as efficiencies in payment processes, 3) the reported high quality of performance discussions in QPRs, and 4) the linkage of organizational performance measures to individual performance appraisals at all levels.

As for conditions that positively predict the use of PI at Treasury, the QCA results align well with the survey and interview results. In an agency whose PMS is sheltered from political influences and resource dependency concerns, and which has an under-developed culture of accountability and empowerment, the role of credible leadership commitment emerges as an indispensable influence to foster federal manager use of performance information. Based on the preponderance of the evidence, Treasury fits this profile well.

PSM Attitudes and Biases. Treasury PSM responses to attitudinal questions indicated they were strong advocates for performance management and believed it offers potential value for agency performance and cost reduction. They see their role as being in charge of "making the agency eat its vegetables," meaning learning new performance management skills that are helpful for the health of the agency but are not especially interesting or desirable for line managers. PSMs noted they feel a sense of fulfillment in helping others accomplish their goals, but struggle to sustain their enthusiasm for institutionalizing performance management, comparing their situation to that of the legendary Sisyphus from Greek mythology. Treasury PSM responses, more than those at all other interviewed agencies, reflected appreciation for the importance of close integration of the performance measurement system in individual manager performance appraisals, performance improvement team activities, and coaching processes. One PSM noted "at Treasury, we've paired the problem finders with the problem solvers."

Case Study #4: Department of Homeland Security (DHS)

Survey and QCA Findings. At DHS, the 2021 survey index scores in table 4.22 for PM (4.40) and PI (4.40) were below the federal mean of agencies surveyed (4.62 and 5.87, respectively). Because the PI index was far below the federal mean of all agencies surveyed, and the lowest of the five agencies with PSMs participating in the interviews, DHS appears to trail

most other federal agencies in routine use of PI for management decisions. Nevertheless, like all federal agencies, DHS has adopted a range of types of performance measures. In contrast with the first three cases, the causal configuration for the institutionalization of performance measures adoption in DHS generated by the QCA was consistent with a sufficient, vice merely necessary, subset relationship. This represents a stronger causal claim than in the cases of HUD, NASA, and Treasury. Both DHS and SBA were identified as ideal-type agencies for a causal combination that included the presence of both a culture of accountability and empowerment and capacity for evaluation of performance. The DHS averages from the 2021 survey align with the QCA findings: the index for C (4.6) was well above the federal average (4.09), and the index for X (4.40) was well above the federal average (3.5). Evidence from the interviews with DHS PSMs at appendix T helps to triangulate and clarify the survey results, and provides deeper insight into C and X and how they act to foster adoption of performance measures in DHS.

Interview Findings. While HUD sees itself as a grants-based agency, and NASA sees itself as a research and development agency, DHS PSMs simply see DHS as a relatively "young" agency that is still maturing its core business models and personnel knowledge of performance matters. PSMs cited agency adolescence as a cross-cutting condition in explaining all aspects of DHS institutionalization of GPRAMA. Like NASA, DHS has a two-tier structure for managing performance. DHS PSMs described the first tier as being "above the water line," to include public-facing performance reports for GPRAMA and other good government laws. The second tier is "below the water line," which includes internal performance evaluation processes which, unlike NASA or HUD, mostly do not include metrics, but do leverage qualitative data from after action reports and mission analyses. Unlike at NASA, there is little integration between the two domains. Employees and managers below the program manager level are generally unaware of

GPRAMA or how performance at their level is related to it. GPRAMA institutionalization in DHS has also deepened over time, but only at the highest levels and only to satisfy OMB mandates, which PSMs noted were the only real tool they had, and greatly appreciated, to get DHS components to comply with GPRAMA reporting mandates.

Evidence from the interviews with DHS PSMs points to an organizational culture with a high degree of manager and organizational empowerment, but middling accountability. Successful managers were empowered through a program of systemic recognition during strategic reviews. As with all previous cases, DHS incentivizes senior executive performance by including strategic plan objectives for which they are responsible in their annual performance reviews.[85] However, at DHS, strong empowerment has a dark side that negatively affects the accountability dimension of organizational culture. Significant empowerment of sub-components through large grants of autonomy to sub-components leaves many DHS components feeling they lack the authority they need to hold their sub-components accountable for their performance against agency priority goals. In addition, PSMs reported that DHS components could deflect criticism for not achieving departmental performance benchmarks by arguing against the validity of the benchmarks themselves based on the significant differences in mission profiles between the DHS components.

The diffusion of authority to DHS components and sub-components appears to be tolerated and thereby reinforced by the strong deference shown to the components by the Secretary of DHS, with the result that the components are sometimes allowed to recuse themselves from GPRAMA requirements. Therefore, while DHS does have a culture that is

[85] PSMs noted that typically 65-70 percent of measures met their target during annual reviews, a level they believed was reasonable because if executives hit their targets 100 percent of the time, then the measures themselves would be suspect.

empowered and accountable based on the indicators used in this study's model for the cultural condition, the interviews revealed an important interaction effect that can occur between the two dimensions. As a result of this mixed assessment, a PSM statement perhaps best sums up performance culture at DHS: "Performance is not high on the list of priorities of most managers." The interview findings therefore provide an important qualification to the survey and the QCA results for the C index.

Organizational capacity to evaluate performance is the second condition present at DHS that is part of the sufficient causal configuration to explain DHS adoption of performance measures. While the 2021 DHS survey index for X was well above the federal mean, a close review of individual survey responses confirmed a majority of DHS respondents affirmed the use of all five types of capacity for evaluation listed in the survey. The interviews mostly confirmed the survey data. For example, DHS PSMs indicated during interviews that back in 2009, DHS had made a major investment into its Internal Controls Program to develop new tools and documentation and establish a process for verification and validation of performance measures which improved data quality that later led to positive results for budgeting processes. DHS components all use a standard database to report their quarterly performance data and to track trends over time. Within components, there are "pockets" of experts engaged in program performance improvement processes.

PSMs confirmed this level of capacity is sufficient for DHS to stay compliant with GPRAMA requirements, but wished to go further to get beyond compliance and implement a system that would satisfy Congress' true intent to make the culture at DHS more performance-oriented. For example, the PSMs noted DHS does not yet do enough process improvement work such as business process re-engineering (BPR), six sigma, or root cause analysis. In response to

interview questions, one PSM suggested that perhaps DHS needs to have a capability to assess its internal performance evaluation capacity. Such comments can be interpreted as a reflection of PSM aspirations to take the DHS PMS to the next level, suggesting it has appropriate capacity to meet current minimum GPRAMA requirements today. Therefore, the interview data align well with the survey and QCA findings for this condition. The strength of this evidence suggests an important finding for the study: investments in capacity for evaluation of performance matter significantly and can make a positive difference for the institutionalization of performance management reforms when coupled with an empowered performance culture.

PSM Attitudes and Biases. Interviewed PSMs at DHS uniformly stated they were supporters of performance management. One PSM self-identified as a "good government geek." The PSMs, however, perceived a difference between their attitudes and those of line managers. PSMs believed line managers generally do not believe implementing GPRAMA is a part of their job requirements, and GPRAMA tasks are not a part of their performance plans. Instead, the PSMs perceive that line managers believe it is the PSMs themselves who are solely responsible for implementing GPRAMA in DHS, because it is a part of their duties. For their part, the PSMs believed it was the responsibility of the Chief Human Capital Officer (CHCO) to create incentives for line managers to institutionalize performance management, but that CHCO had not done so sufficiently. The finger pointing between line managers and PSMs, and between PSMs and CHCO, adds additional evidence to support the findings concerning accountability in DHS culture. Interviewed DHS PSMs also perceived a bias among senior departmental leaders which leads them to focus much more on budgets than performance management, without much awareness of any linkage between them. PSMs perceived that budgeting and program advocacy commanded a far greater share of DHS senior leaders' time and attention than performance

evaluation did. Regarding senior leader perspectives, one quipped: "money is sexy, performance is boring."

Case Study #5: Small Business Administration (SBA)

Survey and QCA Findings. At the SBA, the 2021 survey index scores in table 4.22 for PM (5.00) and PI (6.00) were above the federal mean of agencies surveyed (4.62 and 5.87, respectively). As with DHS, the presence of both a culture of accountability and empowerment and capacity for evaluation of performance were found through the QCA to be consistently sufficient for agency adoption of a wide variety of performance measures. The SBA averages from the 2021 survey partially align with the QCA findings: the index for C (4.00) was just below or virtually at the federal average (4.09), and the index for X (4.00) was above it (3.5). Of note, the SBA index for M in the survey results was exceptionally high (8.00) and was the highest of all agencies surveyed, but was eliminated through Boolean minimization as a condition from the QCA causal configuration, suggesting the absence of M was a prevalent condition in a large number of "easy" (theoretically possible) counterfactual cases. Evidence from the interviews with SBA PSMs at appendix U provides deeper insight into C and X and how they act to foster adoption of performance measures in SBA.

Interview Findings. While DHS PSMs perceived DHS as a "young" agency, SBA PSMs emphasized the *small size* of SBA (PSMs indicated there are only about 3,000 full-time equivalent positions) and its similar role to HUD as a gatekeeper for distributing billions of dollars in federal funds through non-federal partners (i.e., lenders). There was no DHS-like "waterline" in SBA's PMS. In fact, SBA PSMs emphasized the close interactions between the Office of Program Performance Analysis and Evaluation (OPPAE), the SBA focal point for both GPRAMA and Evidence Act institutionalization, and SBA program offices that occur through an

annual cycle of strategy reviews, development of performance plans, and evaluation of programs' progress towards goals. This more centralized approach permits SBA PSMs to better integrate the SBA PMS with other decision-making tools like program evaluation, regulatory analysis, and performance reporting at the SBA headquarters level.

In SBA, PSMs portrayed an organizational culture of strong accountability that started at the very top of the agency, where the Deputy Administrator chairs quarterly performance review meetings with senior leadership. PSMs believed the engagement of the Deputy Administrator creates strong incentives for program officers to be accountable for their decisions and actions. PSMs felt the small size of the SBA staff creates an entrepreneurial orientation in agency culture, suggesting a high degree of individual empowerment. As a result, the OPPAE staff was closely linked to program staff through "informal structures." The interview findings for C, especially with regard to dimensions of manager accountability, are more positive than the PSM survey results. In the PSM survey, SBA PSMs did not agree or strongly agree (nor disagree) with the statements "managers are held accountable for agency strategic goal accomplishment" and "managers are held accountable for goals they are responsible for." Therefore, the survey results may be understated for C. If so, this would only help to confirm the QCA findings for this condition.

Organizational capacity to evaluate performance is the second condition present at SBA that is part of the sufficient causal configuration to explain its adoption of performance measures. SBA PSMs indicated during interviews that performance staff are knowledgeable about procedures to evaluate performance and that SBA has "a good amount of people" to do the work. The OPPAE team combines competencies in performance analysis with programmatic data analysis. PSMs indicated they had no difficulty obtaining requested performance data from

programs. The agency has had data dashboards in place even before it established the capability to conduct program evaluations under the Foundations of Evidence-Based Policymaking Act of 2018, and these dashboards are especially useful for evaluating program-level data. In sum, SBA PSMs indicated PMS capabilities go beyond what is required for GPRAMA, and are part of an even broader suite of services that OPPAE provides to SBA program managers. If there are any capacity constraints on the PMS at SBA, the interview did not discover them. The interview data for SBA thus support the QCA and survey results, and reinforce the findings about the importance of investing in evaluation capacity noted in the DHS case.

PSM Attitudes and Biases. SBA PSMs, like those at DHS, self-identified as professional advocates for performance management. A recurring theme raised by PSMs during the interviews was the high level of collaboration SBA PSMs undertake with all program offices that goes well beyond tasking them to provide performance data needed for GPRAMA reporting. The relationship OPPAE staff have with SBA program offices is enhanced by PSM collaborative attitudes about trying to clearly understand what program offices are trying to accomplish, by helping program offices to find ways to improve their performance, and by offering consultative services to program offices that combine program evaluation, digital dashboards, and performance measurement. One PSM stated "we are genuinely engaged with the programs about how they can improve their activities." All of this points to a high degree of intrinsic motivation and public service motivation of the PSMs and possibly program managers exceeding minimum job task requirements. PSMs indicated program leaders responded positively to this engaging attitude, which built strong and informal relationships between OPPAE staff and program managers.

Synthesis of Case Study Findings

To integrate and synthesize the findings from these five agency case studies based on PSM interviews, the case-oriented predictor-outcome matrix (Miles and Huberman 1994, 214) at table 4.23 presents the researcher's assessment of the degree of presence for each of the system and organization action conditions in the federal agencies which the QCA identified as ideal types for causal configurations either necessary or sufficient for institutionalization of GPRAMA practices (PM and PI).

Table 4.23. Inferential Model Conditions in Federal Agencies Based on PSM Interviews

Paths to GPRAMA practice institutionalization, by federal agency[a]	**Agency Identity**	**P&R**	**C**	**L**	**X**	**M**	**AB**
~P&R~M→PM*							
HUD	"Grant Funder"	O[b]	+	O	+	+[b]	+
NASA	"R&D"	+[b]	++	+++	+	++[b]	++
Treasury	Fiscal Policies/Services[e]	O[b]	+	+++	++	++[b]	+++
*C*X→PM*							
DHS	"Maturing"	+	+[bc]	O[d]	++[b]	+	++
SBA	"Small"	O	+++[b]	+++	+++[b]	+++	+++
~P&R~C*L→PI*							
Treasury	Fiscal Policies/Services[e]	O[b]	+[b]	+++[b]	++	++	+++

Notes:
[a] Field researcher judgment from interview responses
[b] in fsQCA causal configuration solution
[c] mixed indicators: high empowerment but low accountability
[d] Evidence of higher leadership commitment in some component organizations
[e] Source: *Treasury Strategic Plan* FY2018-2022; https://home.treasury.gov/system/files/266/Treasury-Strategic-Plan-FY-2018-2022.pdf
~ "Absence of"
* "and"
+++ Condition fully present
++ Condition partially present or some indicators fully present and others not present
+ Condition minimally present
O Condition not present
P&R Politics and resource dependency
C Accountable and empowered culture

L Credible Leadership Commitment
X Capacity for evaluation
M Mature measurement system
AB Performance System Manager pro-performance management attitudes

Some consistent patterns emerge from this data. First, as noted in the five case analyses, the empirical evidence about the presence or absence of the conditions in the five agencies mostly confirms the results of the QCA. For instance, where the QCA reported the absence of a condition in a causal configuration, data from most PSM interviews confirmed the condition as either nonexistent or minimally present. One exception was NASA and Treasury; PSM interviews revealed greater measurement system maturity than the QCA models did based on the survey data.

Turning to specific conditions, the extent of political influence and resource dependency in all five cases is either non-existent or minimal, offering evidence in favor of the claim that the absence of P&R is a favorable condition for institutionalization of both GPRAMA practices. Comparing tables 4.21 and 4.22, it is striking that the agencies with little or no P&R present and having positive outcomes for PM and PI use were precisely the same ones whose PSMs agreed to be interviewed for this study. Five others included as ideal types in QCA configurations sufficient for adoption of measures where P&R was present (DOC, DOJ, USAID, DOD, and HHS) did not agree to participate in interviews. Is it plausible that their perceptions of the presence of political influence or external interest in their agency's PMS served as a disincentive for their PSMs to participate in interviews? Whatever the case, the evidence on the P&R condition means half of the federal agencies in this study having positive GPRAMA institutionalization outcomes were positively affected by the presence of P&R in combination with C and X, and five were positively affected by the absence of P&R in combination with other conditions. This confirms the findings from the literature review, which strongly suggested

that P&R could enhance or detract from reform institutionalization, depending on the specific characteristics of each agency.

The results concerning the importance of an empowered and accountable culture are more ambiguous than those presented in the quantitative analysis. For SBA and DHS, the two interviewed agencies which shared a sufficient causal configuration for a positive PM outcome, C and X were partially or fully present (the rating for DHS for culture was lowered a level based on the issue of lower accountability indicators within the condition). Apart from the interview evidence, the survey data and QCA results indicated the presence of C combined with the presence of P&R to bring about a positive PM outcome at DOC, DOJ, and USAID. However, the absence of an accountable and empowered culture was a condition in two consistently sufficient causal configurations for routine manager use of performance information: at Treasury, it combined with the absence of P&R and the presence of L, and at DOD it combined with L and X. It appears the presence of C contributes to combinations of conditions supporting the institutionalization of performance measures, but does not facilitate the use of performance information, in contrast to the findings in the quantitative research.

Based on the results of the quantitative research, the qualitative research design did not anticipate that credible leadership commitment would be a condition for configurational analysis of PM adoption, but in three out of five agencies with positive PM outcomes, credible leadership commitment was fully present based on the results of the interviews. In a fourth agency, it did not register at the departmental level but was reported by PSMs as present in one or more component organizations. At Treasury, it was the single present condition in the causal configuration for a positive outcome for PI use. In addition, based on the QCA, leadership commitment was part of a causal configuration leading to PI use in which DoD was an ideal

type. Therefore, evidence from the interviews and the QCA, contrary to the quantitative results, strongly suggests there is indeed a role for demonstrated leadership commitment to advance both forms of GPRAMA institutionalization.

The impact of capacity (X) for evaluation as a causal condition is less clear cut, but the data pattern from the interviews and the QCA reflects agencies that have made substantial investments into evaluation capacity, DHS and SBA, as members of a sufficient causal configuration that includes the presence of C leading to the adoption of performance measures. X also appears in additional causal configurations which lack interview evidence but are in the QCA analysis, in combination with the presence of P&R (DOD, HHS, and USAID) leading to PM adoption, and in combination with L and the absence of C (DOD) leading to PI use. Furthermore, based on the QCA, the absence of X is a condition in a sufficient causal combination for the *non-adoption* of performance measures at DOL and DOT. Integrating the QCA and interview results, the qualitative study confirms the findings of the quantitative study about the importance of making investments into capacity for evaluation as a causal condition supporting both forms of GPRAMA institutionalization.

Interview evidence on the maturity of the performance measurement system (M) shows no obvious pattern across interview cases. When combined with the results of the QCA in table 4.21, the evidence from the interviews may permit the postulation that the absence of M is a contributing condition in a necessary causal configuration leading to performance measure adoption (HUD, NASA, and Treasury) and the presence of M is a condition in a sufficient causal configuration leading to the non-adoption of performance measures (DOL and DOT). The qualitative research therefore finds that the absence of M is a favorable condition for the institutionalization of performance measures adoption. As previously noted, a mature

measurement system may have little need for additional new measures. Therefore, this finding, while it contradicts some existing empirical research in the performance management literature as noted in chapter 2, makes intuitive sense.

Regarding PSM attitudes and biases, there was only minor differentiation between the six who interviewed; all except one demonstrated convincing personal support for performance management as a function, even independent of GPRAMA requirements, and in most cases their interview responses indicated a high level of intrinsic motivation and a desire to see performance management serve as a useful management tool for agency managers, and not just be a compliance exercise. Across cases, there was evidence that PSM attitudes were strongly affected by the demonstrated credible commitment of senior agency leaders. Senior leaders who showed genuine interest and engagement in their agency's PMS, and demanded senior managers and programs participate in GPRAMA routines, generally inspired their PSMs. Those who saw GPRAMA as a burden or a compliance exercise tended to demotivate their PSMs. Turbulent and short-lived tenures of political appointee leaders likewise created uncertainty and reduced PSM and line manager incentives to deepen the institutionalization of GPRAMA practices.

Somewhat surprisingly, there was at best a weak association between agency decisions concerning resourcing of performance management capacity and PSM attitudes about GPRAMA or their work. PSMs with higher levels of intrinsic motivation acknowledged resource shortfalls in their PMSs, but tended to also point out low-cost or no-cost innovations their agency had undertaken, such as monthly (vice quarterly) Baseline Performance Reviews at NASA. There is, however, insufficient evidence from the interviews to make the claim that poor PSM attitudes contribute to the weak institutionalization of GPRAMA or that positive PSM attitudes are instrumental in successful institutionalization. PSMs play a facilitative and educative role in

GPRAMA institutionalization in their agencies, but critical decisions about the selection of performance measures and management decisions to use performance information are generally reserved for agency executive leadership and program managers, who were outside of the scope of data collection efforts, one of several limitations in the study.

Chapter 5. How Agency Identity and Bureaucratic Type Affect GPRAMA Institutionalization

The second research question presented in chapter 1 asks: "How does the nature of an agency's identity according to its bureaucratic type, specifically Production, Procedural, Craft, or Coping (J. Q. Wilson 1989) affect its adoption of performance measures and use of performance information for management decisions?" Chapter 2 briefly described the taxonomy of bureaucratic types (figure 2.2) proposed by James Q. Wilson (1989) in his canonical text, *Bureaucracy.*[86] Wilson's taxonomy places public agencies fall into four ideal types: Production, Procedural, Craft, and Coping, depending on the level of observability of their outputs and outcomes. Agency membership in a bureaucratic type set theoretically implies conditions favorable for the institutionalization of performance management practices exist based on the relative level of observability of an agency's outputs and outcomes. The positioning of agencies within categories is, of course, relative. Few agencies represent examples of ideal types, but some agencies come closer to the ideal than others and can serve as the basis for comparisons.

Wilson (1989) argues when outputs and outcomes are easily observable, all else being equal, managers' tasks are simplified and they are better positioned to develop systems for achieving efficient outcomes (160). Furthermore, he argues managers will tend to allocate their attention, intentionally or subconsciously, towards achieving and measuring outcomes which are more observable, and neglect those which are not (161). For Production agencies, this tends to lead agencies to embrace a "stat game" approach for their performance measurement systems which can create incentives for employees to try to manipulate the game in their favor (161-162).

[86] Wilson (1989) gave due acknowledgement to Henry Mintzberg's ideas in *The Structuring of Organizations* (1979) as a source of inspiration for his taxonomy (159).

In Procedural agencies, Wilson argues managers become "means oriented" and focus more attention on how employees perform their work than on program outcomes; adhering to procedural constraints becomes a performance measure in itself (164). In Craft agencies, management can easily observe outcomes, but not outputs. This drives them to be more goal-oriented and less means oriented, which implies far less oversight of employee activities (166). In the absence of centralized controls, effective management of Craft agencies requires those agencies' identity include a strong ethos of skills-based professionalism based on formal training and job tradecraft (167). Wilson argues managers in Coping agencies struggle to define what constitutes "good work" because there is no easy way to measure either outcomes or outputs, which leaves them little to measure performance of their agency beyond constituent complaints or crises (169). This fact creates a professional dilemma for managers in Coping agencies who may wish to make use of performance information, a dilemma that likely results in decisions to make little use of it:

> And so managers, depending on their personal style, cope with the complaints as best they can. In doing so, they must strike a delicate balance: if they take the complaint seriously, the operator will feel that he or she has not been 'backed up'; if they take the complaint lightly, the citizen will feel that the agency is 'insensitive.' Teachers do not like principals who fail to back them up in conflicts with pupils or parents. Police officers do not like captains who fail to back them up in conflicts with citizens and lawyers (J. Q. Wilson 1989, 170).

These varying management preferences, in turn, theoretically influence management decisions as to whether and how to institutionalize particular practices to support GPRAMA in their agencies, resulting in different agency responses to reform mandates. Chapter three applied Wilson's taxonomy to analyze variance in agency responsiveness to GAO recommendations (table 3.5) and federal manager perceptions of performance information use in their agencies (figure 3.2) during the Obama and early Trump administrations and found circumstantial

evidence to support this claim based on historical evidence between bureaucratic type and the studied variables, especially concerning Coping agencies.

In this study, the concept of agency identity takes advantage of Wilson's theoretical insights to operationalize a proxy for the abstract concept of agency identity in two ways. First, agency identity may generate a strong, weak or neutral latent performance orientation in federal agencies relative to that of other agencies which have different agency identities. The theoretical expectation is that the organizational identity of every agency, based on its legal charter and policy mandates, determines the nature of all of its activities and programs, and the degree of outcome and output measurability of these activities and programs varies, sometimes greatly. Simply put, if Wilson (1989) is correct, then agencies with programs whose outputs and outcomes are easily observable and measurable are more likely to try to measure them and use performance information about them to improve their performance, or at least the appearance of performance. They therefore may be considered as possessing a stronger performance orientation, the characteristic of agency identity of interest in this study. Agency performance orientation is not easily directly measurable, but its effects, if any, should be discoverable through the measurement of the effects of agency identity on GPRAMA institutionalization outcomes. Chapter 2 offered two preliminary hypotheses that this chapter tests:

$H_{9a.}$ A significant relationship exists between an agency's identity and its adoption of performance measures.

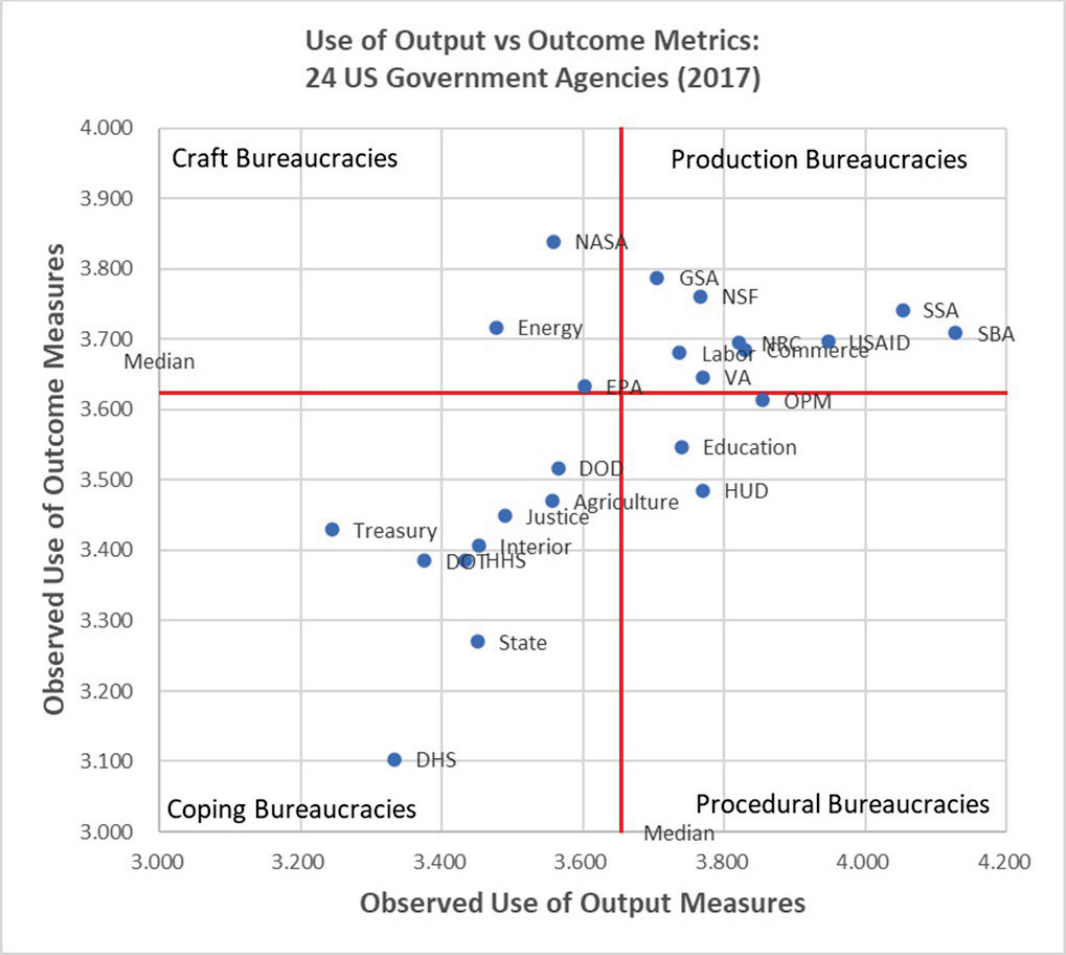
$H_{9b.}$ A significant relationship exists between an agency's identity and its use of performance information.

According to these hypotheses, I expected to observe evidence of significant variation between agencies in terms of their adoption of performance measures and use of performance information

based on their agency identity alone. The purpose of these hypotheses was merely to test for the existence of the postulated effects of agency identity, not to make a claim regarding the direction of the relationship, a purpose reserved for the second investigation of agency identity.

Second, to test Wilson's claims about the behavior of different bureaucratic types towards performance management institutionalization in the context of the U.S. Government, I measured output and outcome observability of the 24 federal agencies separately using perceptual evidence from the 2017 GAO survey of federal managers and used the results to classify each of the 24 CFO Act agencies into one of the four Wilson bureaucratic types based on its relative standing vis-à-vis the other 23 agencies. Specifically, federal manager responses to two GAO 2017 survey questions on their agency's use of outcome and output measures, averaged by agency clusters, served as indices of agency outcome and output observability. I used a scatterplot diagram with agency labels and quadrant borders based on median index values for both indices to cluster agencies into the four Wilson bureaucratic types. The further from the universal median an agency is, as measured by the combined variance of its output and outcome distance from the universal median, the closer it is to being an ideal type for the Wilson category. Therefore, NASA is the most ideal Craft agency, the SBA is the most ideal Production bureaucracy, DHS has no close rival to being the most ideal Coping agency, and HUD is the most ideal Procedural agency. Figure 5.1. presents the results of this classification strategy.

Figure 5.1. Classification of Federal Agencies by Wilson Bureaucratic Type



Source: Author's own design using federal manager responses to questions 5a and 5e in the 2017 GAO survey of federal managers.

In order to test Wilson's hypotheses using empirical evidence from U.S. Government agencies, chapter 2 advanced four hypotheses:

$H_{10a.}$ Production agencies are more likely to adopt output measures than Craft and Coping agencies are, and are more likely to adopt outcome measures than Procedural and Coping agencies are. They are more likely to use performance information than all other bureaucratic types.

$H_{10b.}$ Procedural agencies are more likely to adopt output measures than Craft and Coping agencies, and are less likely to adopt outcome measures than Craft and Production agencies. They are more likely to use performance information than Coping agencies.

$H_{10c.}$ Craft agencies are more likely to adopt outcome measures than Procedural and Coping agencies, and are less likely to adopt output measures than Production and Procedural agencies. They are more likely to use performance information than Coping agencies.

$H_{10d.}$ Coping agencies are less likely to adopt performance measures of any kind and to use performance information for any purpose than Production, Procedural, and Craft agencies are.

Quantitative Research Design

Introduction

The research strategy in table 4.1 called for the use of quantitative methods to study the effects of agency identity and bureaucratic type on the two primary indicators of GPRAMA institutionalization in this study, PM and PI. The research design to answer research question 2 employs in part the same HLM quantitative models (figure 4.7) and estimation methods applied to answer research question 1, with the exception that random effects parameters are predicted sequentially after model fitting and estimation, and assume independent covariances to enable separate variances for each random effect. Otherwise, it shares the same data source and sample (federal manager responses to the 2017 GAO survey), the same data collection and preparation procedures, the same tests of OLS assumption violations and the same strategies to mitigate potential threats to validity. However, the construction of the variable for agency identity in these models has not yet been explained. The next section addresses this.

Construction of Variables

The 2017 GAO survey dataset provided an agency stratification code as a nominal variable between 1 and 24. This study leveraged that nominal variable with 24 categories as the second-level group variable in the mixed effects models 1 and 2 to measure the random intercept at the agency level.[87] This approach ensured all manager observations were appropriately clustered by agency, and so that the random effects of agency identity on the dependent variables could be separately predicted and compared across agencies. A stronger random effect would represent a greater positive or negative association between the agency's identity and the dependent variables.

In addition, the study created another nominal variable, Wilson bureaucratic type, with four categories, to reflect the four types of bureaucracies in Wilson's taxonomy of bureaucracies. This variable was used not only to estimate the fixed effects of bureaucratic type on the two primary indices of GPRAMA institutionalization, PM and PI, but also to estimate the fixed effects of bureaucratic type on each of the 14 GPRAMA practices in this study.

Units of Observation and Measurement Models

As with models 1 and 2, the unit of observation remains the federal managers who responded to the 2017 GAO federal manager survey. To evaluate the effects of agency identity on federal agency adoption of performance measures and use of performance information, models 1 and 2 are recycled here, but additional analysis on agency identity random effects is conducted to assess cross-agency variance. Bell and Jones (2015) argue a random effects model is suitable for studying research questions in contexts with hierarchical structures, and is normally superior to purely fixed effects models due to its flexibility, generalizability, ability to

[87] Details on Stata procedures used to fit mixed effects models is available at: https://www.stata.com/manuals/me.pdf (especially pp. 503-516).

model the causes of endogeneity, and, for purposes of this study, its "…ability to model context, including variables that are only measured at the higher level" (1-2). Use of a second level variable in mixed effects models can help to reduce omitted variable bias at the observation level by accounting for multiple sources of variation among the fixed effects (Bell, Fairbrother, and Jones 2019, 1059). This fact also offers a degree of added confidence in the validity and significance of the fixed effects coefficients for the system action and organization action situation conditions presented in chapter 4.

Use of fixed effects modeling is also a potentially useful approach for studying the effects of agency identity. The Hausman specification test, however, revealed a significant difference in coefficients between fixed and random effects variants of model 1 (chi^2=24.87, p=.0008), but no systematic difference for model 2 (Chi^2 = 2.97, P=.8874). Therefore, use of a random effects approach for model 1 makes sense, but for model 2 it is not required. It is nevertheless desirable because, according to Schofer (2007), the random effects model is more efficient than the fixed effects model when there is no significant difference in the coefficients of the random and fixed effects (n.p.). In addition, through use of simulations, Bell, Fairbrother, and Jones (2019) found that ". . . estimated standard errors are anti-conservative when random slope variation exists but a model does not allow for it" (1052). Therefore, to conservatively predict how individual agencies vary across the government in terms of their relative identity effects on PM and PI, the best linear unbiased predictions (BLUPs) of random effects of agency identity are calculated using post-estimation techniques from models 1 and 2, immediately after fitting the regression models. No new variables were constructed for this analysis. The same random effects parameters for federal agencies were included in HLM models 1 and 2 at level two.

Unfortunately, because random effects at the second level of a mixed effects model can be predicted only after a model is fit, they must be predicted without the benefit of MI using complete case analysis. Therefore, post-regression estimates of random effects of agency identity are based on a smaller sample size than the fixed effects estimates using MI (model 1 used 543 observations, model 2 used 455). Estimating the random effect intercepts and slopes for agency identity using an imputed single level model is a feasible option, but could pose a threat to validity. Van Buuren (2018) argues ignoring agency clustering risks underestimating intra-class correlations and biasing the regression weights, and concludes that "conducting multiple imputation with the wrong model (e.g. single level methods) can be more hazardous than listwise deletion" (202-203). Because of the previously-indicated advantages of using a bi-level mixed effects model, allowing a smaller sample size to estimate agency identity effects is a conscious tradeoff, thereby giving priority to appropriate method over additional (imputed) sample size.

To study the effects of bureaucratic type on specific practices of GPRAMA institutionalization, model 3 at figure 4.8 was used as a starting point for model 4. In model 4, bureaucratic type was fitted as a single independent categorical variable, replacing all other predictors to ensure its full effects could be discovered and estimated. A limitation of this approach is that bureaucratic type cannot be modeled as a second level random effect using a partial proportional odds model as it can be using a mixed effects model, so the effect of bureaucratic type on the adoption of specific performance measures and uses of performance information must be modeled instead as an observation-level fixed effect. This allows within-agency effects for model 4 to be estimated and then separately compared. However, the

interpretation of these estimates will not be extendable beyond the population of 24 CFO agencies as they could be using a random effects model.[88]

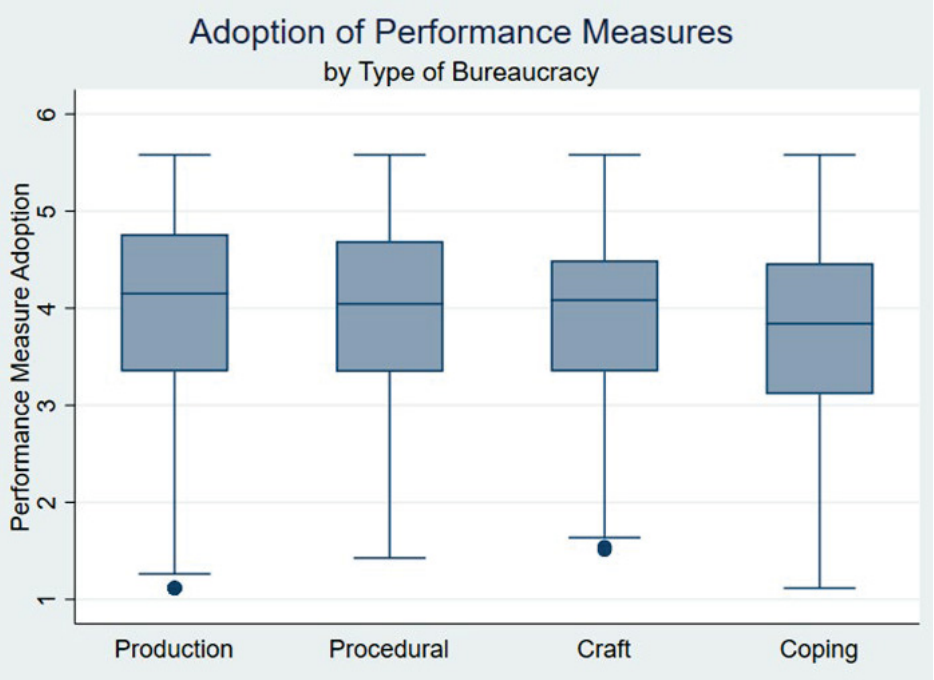
Quantitative Findings

Summary of Descriptive Data from 2017 GAO Dataset

Descriptive box and whisker plots using data drawn directly from the 2017 GAO survey dataset (figure 5.2 and figure 5.5) provide a preliminary look into differences between Wilson bureaucratic types concerning federal manager perceptions about their agencies' adoption of performance measures and use of performance information. Note especially the range and variable placement of the box containing the upper and lower quartiles and median, which declines on average from Production to Coping agency types, an indication of variance in performance orientation between agency types. Here, Production, Procedural, and Craft agencies are all higher on average than Coping agencies, as Wilson's theory would suggest. The differences are not so large, however, as to provide conclusive evidence of a significant difference, but certainly are adequate to justify more sophisticated analysis.

Figure 5.2. Descriptive Data on Performance Measure Adoption, by Wilson Bureaucratic Type

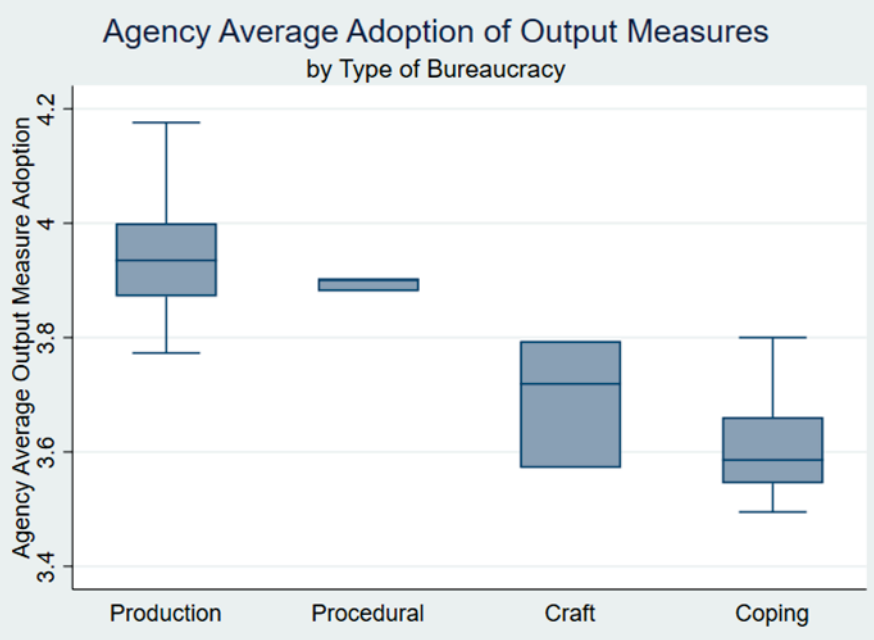
[88] Given how much of the US Government these 24 departments and agencies represent (over 90 percent of employees), this should not be a major concern.



Notes:
Source: 2017 GAO Survey of Federal Managers
Shows raw (non-standardized and non-transformed) data from GAO dataset.

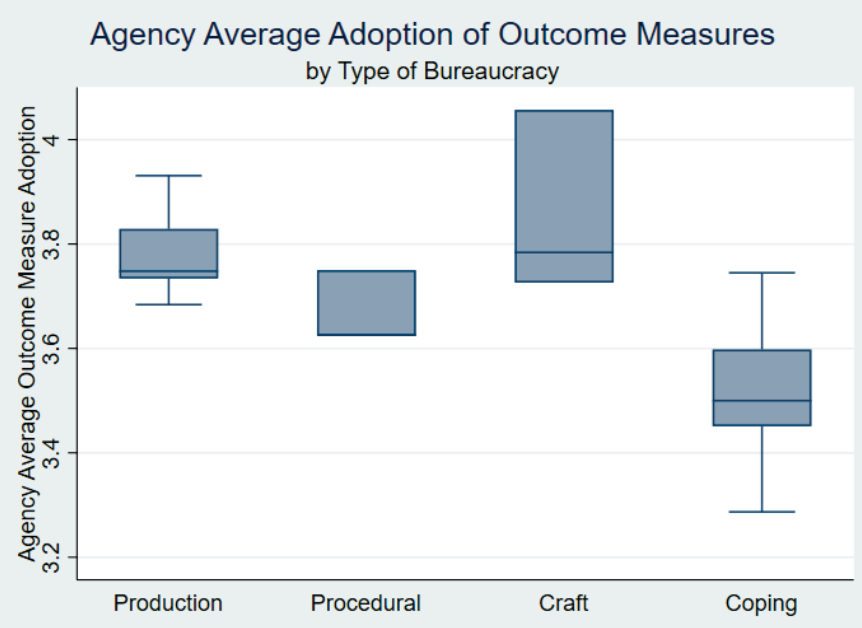
These data, however, encompass all categories of performance measures in the study, and thereby obscure the theoretically significant question of how responsive each type of agency is towards adoption of output and outcome measures. Figures 5.3 and 5.4 provide a preliminary look into differences between Wilson bureaucratic types concerning federal manager perceptions about their agencies' adoption of output and outcome measures. Here, the evidence is plainer. Production and Procedural agencies have average indices of adoption of output measures that are substantially higher than those of Craft and Coping agencies, and Craft and Production agencies have average indices of adoption of outcome measures that are higher than Coping and Procedural agencies. This data provides statistical comparisons that more clearly reflect the distribution of agencies along the output and outcome measures adoption spectrums in figure 5.1.

Figure 5.3. Descriptive Data on Output Measure Adoption, by Wilson Bureaucratic Type



Notes:
Source: 2017 GAO Survey of Federal Managers
Shows raw (non-standardized and non-transformed) data from GAO dataset.

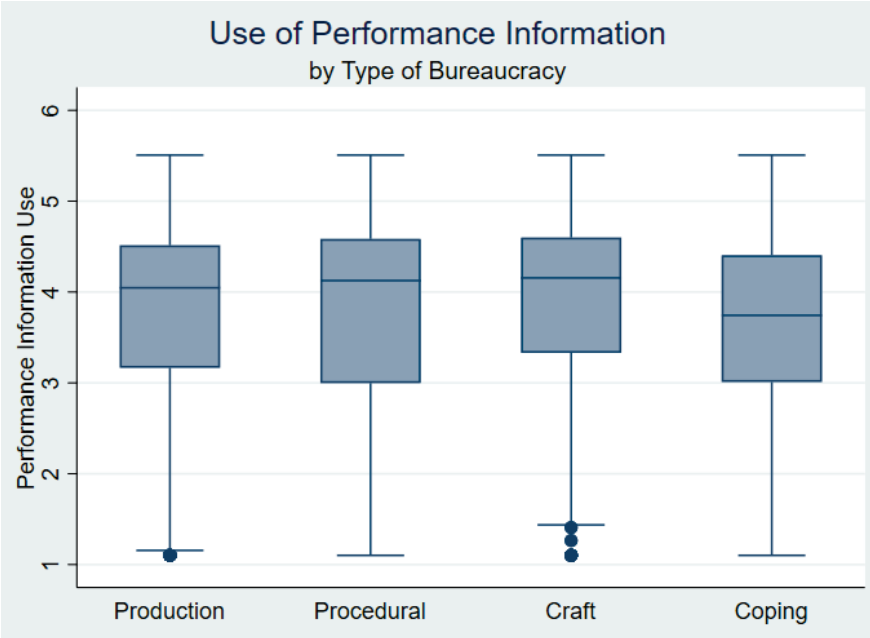
Figure 5.4 Descriptive Data on Outcome Measure Adoption, by Wilson Bureaucratic Type



Notes:
Source: 2017 GAO Survey of Federal Managers
Shows raw (non-standardized and non-transformed) data from GAO dataset.

Turning to the other institutionalization indicator, use of performance information, the descriptive data are more ambiguous. The only substantial difference is between Coping agencies and all other types. As a cluster, Craft agencies demonstrated the highest overall tendency to use performance information, even higher than Production agencies. This suggests some aspects of craft agencies' identity are strongly reliant on using performance data for management decisions.

Figure 5.5. Descriptive Data from 2017 GAO Survey on Performance Information Use, by Wilson Bureaucratic Type



Notes:
Source: 2017 GAO Survey of Federal Managers
Shows raw (non-standardized and non-transformed) data from GAO dataset.

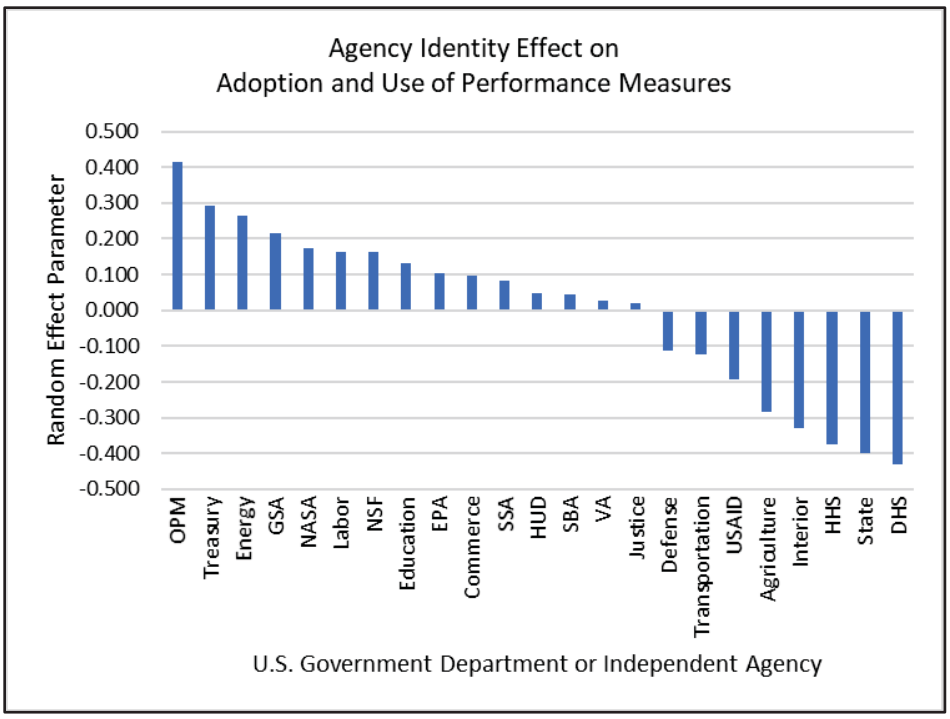
Effects of Agency Identity on GPRAMA Institutionalization Indices

Descriptive data is helpful to gain a broad understanding of trends and to suggest additional avenues for investigation, but is not able to isolate and estimate the association of agency identity or bureaucratic type on any of these institutionalization outcomes while controlling for the effects of other model conditions. Any of these conditions may be largely

responsible for the results in figures 5.2 through 5.5. For this, we need to employ regression analysis. We begin by analyzing between-agency variance in the effects of agency identity on PM and PI, as noted, by fitting HLM models 1 and 2 and then predicting the random effects post-estimation. The random effect parameters for each agency represent variance from the global mean of all agencies' random effects. The presence of significant variation between agencies' random effects parameters for both models 1 and 2 is shown in figures 5.6 and 5.7.[89] The results appear to support hypotheses H_{9a} and H_{9b}. I find that a significant relationship exists between an agency's identity and its adoption of performance measures and between an agency's identity and its use of performance information, although the magnitude of agency identity effects on agency performance information use is consistently less than they are on agency adoption of performance measures.

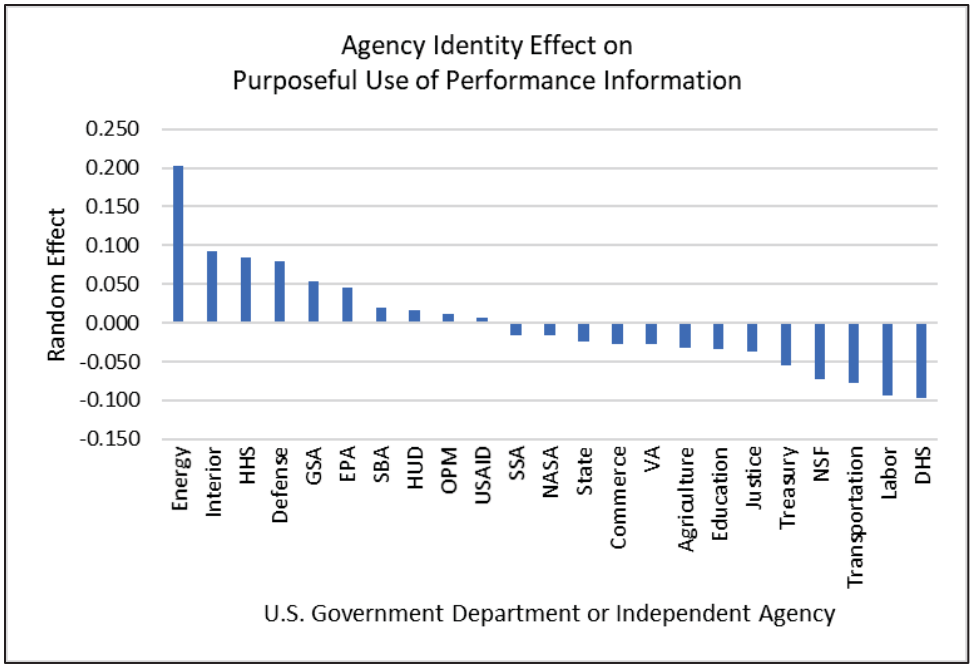
Figure 5.6. Random Effect of Agency Identity on Adoption of Performance Measures

[89] Unlike fixed effects, it is not possible to calculate degrees of freedom or test for statistical significance of random effects. Random effects parameters do not have asymptotic distributions or calculable degrees of freedom which are required to generate t-test statistics, according to the Social Science Research Cooperative at the University of Wisconsin, which states that "one source of the complexity is a penalty factor (shrinkage) which is applied to the random effects in the calculation of the likelihood (or restricted likelihood) function the model is optimized to. This results in distributions which are no longer chi squared or F. This penalty factor also complicates determining the degrees of freedom to associate with the estimate of a random effect." (https://www.ssc.wisc.edu/sccc/pubs/MM/MM_TestEffects.html).



An agency's identity therefore plays an important role in the selection of performance measures, but a less significant role in management decisions to use performance information. The data suggest, for instance, that some characteristic of OPM's identity strongly explains its adoption of performance measures, and that some feature of DHS's identity strongly detracts from its adoption of performance measures. From the case study evidence on DHS in chapter 4 discovered through the 2021 survey and interviews with the author, its shared identity as a "young" agency that is still maturing its core business models is a probable cause for this result. Indeed, DHS PSMs interviewed uniformly cited agency adolescence as a cross-cutting condition in explaining all aspects of DHS institutionalization of GPRAMA. In this sense, the maturity of an agency's business functions may be considered an element of its agency identity.

Figure 5.7. Random Effect of Agency Identity on Use of Performance Information



Observable patterns hidden within these agency identity random effects data, however, are anything but random. A close analysis of the federal agencies in this study and their performance orientations shows different types of agencies, based on their performance orientation, are more positively or negatively associated with the institutionalization of GPRAMA practices.

Cluster analysis of agency random effects in table 5.1 plausibly supports Wilson's theoretical predictions about the probability of adoption and use of PM based on bureaucratic type: Mean random effect scores by Wilson bureaucracy type are all positive for Production, Procedural, and Craft agencies, but Coping agency mean random effects are strongly negative, in line with Wilson's claims. Seven out of eight Production agencies have positive random effects, and all Procedural and Craft agencies have positive effects. Seven out of nine Coping agencies have negative effects. This cluster pattern strongly suggests that the type of bureaucracy an agency is has an influence on its likelihood of adopting performance measures, although Wilson's theory would suggest Production agencies, not Procedural agencies, should have the

highest overall mean agency random effect. Table 5.1 shows the results of figure 5.6 sorted by Wilson bureaucratic type and magnitude of effect from high to low within each type cluster.

Table 5.1. Random Effects of Agency Identity on Adoption of Performance Measures

Wilson Type	Agency	Agency Name	Effect	Type Mean
Production	10	GSA	0.215	0.075
Production	6	Labor	0.165	0.075
Production	15	NSF	0.163	0.075
Production	3	Commerce	0.096	0.075
Production	21	SSA	0.083	0.075
Production	19	SBA	0.046	0.075
Production	24	VA	0.028	0.075
Production	2	USAID	-0.193	0.075
Procedural	18	OPM	0.414	0.198
Procedural	8	Education	0.133	0.198
Procedural	13	HUD	0.047	0.198
Craft	7	Energy	0.264	0.181
Craft	16	NASA	0.174	0.181
Craft	9	EPA	0.104	0.181
Coping	23	Treasury	0.294	-0.193
Coping	5	Justice	0.019	-0.193
Coping	4	Defense	-0.112	-0.193
Coping	22	Transportation	-0.123	-0.193
Coping	1	Agriculture	-0.285	-0.193
Coping	14	Interior	-0.330	-0.193
Coping	11	HHS	-0.375	-0.193
Coping	20	State	-0.400	-0.193
Coping	12	DHS	-0.429	-0.193

Notes:
The NRC, a Production-type agency, was excluded due to GAO survey instructions for managers to skip series 17 questions, a component of the goal clarity predictor in model 1.

Similar to the above results, I note that agency random effects on the use of performance information also vary, although the effect magnitudes are about half of those for performance measures adoption. For agency identity effects on performance information use, there is no

consistent or discernable pattern of agency random effects within Wilson bureaucratic type clusters. Only three of eight Production agencies have positive random effects, two of three Procedural agencies do, and two of three Craft agencies do. Three of eight Coping agencies also have positive effects. Table 5.2 shows the results of figure 5.7 sorted by Wilson bureaucratic type and magnitude of effect within type cluster.

Table 5.2. Random Effects of Agency Identity on Use of Performance Information

Wilson Type	Agency	Agency Name	Effect	Type Mean
Production	10	GSA	0.054	-0.02
Production	19	SBA	0.02	-0.02
Production	2	USAID	0.006	-0.02
Production	21	SSA	-0.016	-0.02
Production	3	Commerce	-0.028	-0.02
Production	24	VA	-0.028	-0.02
Production	15	NSF	-0.073	-0.02
Production	6	Labor	-0.094	-0.02
Procedural	13	HUD	0.017	-0.002
Procedural	18	OPM	0.011	-0.002
Procedural	8	Education	-0.034	-0.002
Craft	7	Energy	0.203	0.077
Craft	9	EPA	0.045	0.077
Craft	16	NASA	-0.016	0.077
Coping	14	Interior	0.092	-0.007
Coping	11	HHS	0.085	-0.007
Coping	4	Defense	0.079	-0.007
Coping	20	State	-0.024	-0.007
Coping	1	Agriculture	-0.033	-0.007
Coping	5	Justice	-0.037	-0.007
Coping	23	Treasury	-0.055	-0.007
Coping	22	Transportation	-0.077	-0.007
Coping	12	DHS	-0.097	-0.007

Notes:
The NRC, a Production-type agency, was excluded due to GAO survey instructions for managers to skip series 17 questions, a component of the goal clarity predictor in model 1.

Effects of Bureaucratic Type on GPRAMA Institutionalization Indices

To check the robustness of the findings of the bureaucratic type cluster analysis of the agency identity random effects, and to measure covariances of bureaucratic type with the two GPRAMA measures of institutionalization, I fitted single-level fixed effects OLS regression models using robust standard errors and agency clusters with PM and PI as the dependent variables and the nominal variable Wilson bureaucratic type as a predictor. The results in table 5.3 partially confirm portions of the agency identity random effects cluster analysis. With Production bureaucracies as the base case for each check, the regression coefficients for the other three bureaucratic types generally comport with the cluster analysis results as well as Wilson's claims regarding the effects of different bureaucratic types on agency adoption of performance measures (the coefficient for Craft bureaucracies was not significant). The magnitude of the effect on measures adoption is small, however: r^2 was only .0099. I find that bureaucratic type significantly predicts adoption of performance measures at the .01 level ($F=5.32$, $p>F=.0062$), but by itself explains only about 1 percent of the variation in agency adoption of performance measures.

For the effect of bureaucratic types on the use of performance information, the results in table 5.3 also reflect the higher level of ambiguity noted in the cluster analysis and in the descriptive data. Only one coefficient (for Craft agencies) was positive and significant, but only at the .05 level and based on use of robust standard errors. This result validates the previous observation from the descriptive data about the possibility that one or more characteristics of Craft agencies may lead their managers to be more routine users of performance information for management decisions than in other types of agencies. Nevertheless, the coefficients for the other types roughly comport with the direction of Wilson's theory. Procedural agencies were slightly lower than Production types, and Coping agencies were, as with the descriptive and

random effects cluster findings, the lowest of all types. The magnitude of the effect of bureaucratic type on use of performance information is very small, however: r^2 was only .0028. I find that bureaucratic type significantly predicts performance information use at the .05 level (F=3.10, p>=.047), but by itself explains less than half a percent of the variation in agency use of performance information.

Table 5.3. Effects of Wilson Bureaucratic Type on GPRAMA Institutionalization Indices

		Wilson Bureaucratic Type		
Dependent Variables	**N**	**Procedural**	**Craft**	**Coping**
Adoption of Performance Measures	2,151	-.124(.054)*	-.069(.143)	-.249(.070)**
Uses of Performance Information	1,643	-.017(.044)	.187(.077)*	-.052(.125)

Notes:
Fixed Effects OLS Model with sample weights and agency clusters
Production Bureaucracies are base case
** p<.01, * p<.05, + p<.10.
Robust standard errors in parentheses.

Results in figures 5.3 and 5.4 suggest a significant difference between bureaucratic types exists in terms of their adoption of output and outcome measures, but descriptive evidence alone cannot confirm that. Results in tables 5.1-5.3 measure the effects of agency identity and bureaucratic type on the concepts of PM and PI, both of which are highly aggregated operationalizations of a large number of related performance management practices (PM=5 routines, PI=9 routines). Identifying patterns of agency institutionalization of GPRAMA routines at that level of abstraction is designed to inform theoretical debates in the governance and public performance management literature. Aggregating so many related routines, however, obscures significant and interesting differences between agencies concerning what kind of measures they adopt and what ways their managers choose to use performance information in planning and operations.

Effects of Agency Identity on the Institutionalization of 14 GPRAMA Practices

Do Production and Craft bureaucracies favor the use of outcome measures, as Wilson (1989) suggested? Do Production and Procedural bureaucracies favor the use of output measures? Are Coping bureaucracies systemically challenged by their very nature in their adoption of outcome, output, and perhaps other kinds of measures as well? Are some types of bureaucracies more or less likely to make greater use of performance information to set performance goals, manage contracts, or take corrective actions? To answer these and similar questions concerning the other kinds of performance measures and uses of performance information, model 4 adapts the fixed effects measurement strategy in model 3 by replacing all predictors with agency identity as a nominal categorical variable, while retaining the same model scheme of agency clustering, partial proportional odds, and dependent variable threshold using the GOLOGIT2 add-on function in Stata 15.

Results for the role of agency identity reveal it is both positively and negatively correlated with agency institutionalization of different kinds of performance measures and uses of performance information based on the agency's bureaucratic type. In appendix F and appendix G, a fairly consistent pattern of agency associations can be seen that shows the fixed (within-group) effects of agency identity on all 14 GPRAMA practices using model 4. In these results, the EPA was selected as the base case due to being the agency with the closest proximity to the universal median of all agencies for observability of output and outcome measures. This approach aids interpretability of odds ratio coefficients of the other 23 agencies. Odds ratios in appendices F and G that are greater than 1 indicate an agency is more likely to institutionalize the routine, and odds ratios less than 1 suggest an agency is less likely to institutionalize it. Only statistically significant coefficients are displayed.

The results in appendix F support hypotheses 10a and 10d but are inconclusive for 10b and 10c. In the case of Production agencies, there are 23 instances in which agency identity significantly improved the likelihood one or more of the five measures would be adopted, and no instances in which it reduced it. For three of nine Production agencies, their identity made it more likely the agency would adopt output measures, but the same was true for only one agency regarding adopting outcome measures. For the other three kinds of measures, being a Production agency was a fairly strong predictor of adoption: efficiency measures (100 percent of agencies), customer services measures (55 percent), and quality measures (55 percent).

Procedural agencies and Craft agencies had far fewer significant instances to evaluate than Production agencies did, so it is difficult to draw firm conclusions about them based on the limited evidence available. In terms of Wilson's predictions, the evidence that is available here is mixed. Contrary to Wilson's predictions, none of the three Procedural agencies showed a significant association with agency adoption of output measures. Being a Procedural agency did improve the chances of adoption of efficiency measures (100 percent of agencies), customer service measures (67 percent) and quality measures (33 percent). As Wilson predicted, it did not increase the likelihood of adoption of outcome measures.

For the nine Coping agencies, the results strongly support Wilson's claims. Only one agency was significantly associated with the adoption of output measures (State) and its odds ratio was only .626, suggesting the association was strongly negative. For outcome measures, four out of nine agencies had significant coefficients, and all were less than one, with DHS the lowest at .534. In all of these cases, the Coping nature of the agency's identity reduced the likelihood the agency would adopt output and outcome measures. Finally, Coping agencies demonstrated a more favorable relationship with the adoption of efficiency measures (67 percent

of agencies, all positive), customer service measures (22 percent, all positive), and quality measures (22 percent, all positive) than they did with either output or outcome measures.

The results in appendix G offer less conclusive evidence about the effects of agency identity on PI use, but hypotheses 10a, 10c, and 10d seem to be supported. For Production agencies, out of 18 instances of a significant effect across the nine forms of PI use, 14 show an increased likelihood of PI use (coefficient >1). As above, there is less available evidence for Procedural and Craft agencies. There is only a single significant instance for Procedural agencies, but six for Craft agencies (of which five showed an increased likelihood). Again, the evidence for the effect of Coping agency identity is more convincing, with 25 significant instances, all of which were less than one. The evidence therefore suggests a finding that Production agency identity significantly improves the odds of PI use, and Coping agency identity significantly reduces the odds of PI use. Some evidence also points to a systemic tendency towards PI use due to an agency's identity as a Craft bureaucracy, even more so than for Production agencies, a surprising finding

Effects of Bureaucratic Type on the Institutionalization of 14 GPRAMA Practices

These results are conceptually similar to the descriptive results presented earlier which clustered the raw data from the GAO dataset by agency. The evidence in appendix F and appendix G points to the possibility that positive and negative relationships exist between bureaucratic type and the institutionalization of different GPRAMA practices. To test the robustness of these findings from the analysis of clusters of significant agency-level regression coefficients, I replaced the 24-category nominal agency identity predictor with the four-category nominal Wilson bureaucratic type variable in model 4 to measure the likelihood that different bureaucratic types would institutionalize the fourteen GPRAMA practices in this study and to

investigate if there are any systemic differences in that likelihood between the bureaucratic types. The results are in table 5.4. As with table 5.3, Production agencies are the base case (equal to 1 for all practices). The other three bureaucratic types are compared with them. Here, coefficients are expressed as the odds that bureaucratic type will increase federal manager agreement about the institutionalization of the specified practice to the "great extent" level from the "moderate extent" level in the 2017 GAO survey of federal managers.

Table 5.4. Effects of Wilson Bureaucratic Type on GPRAMA Practices

		Wilson Bureaucratic Type		
Dependent Variables	**N**	**Procedural**	**Craft**	**Coping**
Adoption of Performance Measures				
Output Measures	2,333	.880(.074)	.627(.063)**	.527(.047)**
Efficiency Measures	2,344	1.121(.172)	.621(.158)+	.666(.092)**
Customer Service Measures	2,296	.906(.190)	.986(.308)	.791(.153)
Quality Measures	2,336	.809(.186)	.963(.260)	.715(.115)*
Outcome Measures	2,342	.810(.071)*	1.092(.178)	.613 (.055)**
Uses of Performance Information				
Developing Program Strategy	2,535	.899(.089)	.914(.101)	.717(.099)*
Setting Program Priorities	2,543	1.017(.088)	1.024(.088)	.707(.075)**
Allocating Resources	2,507	.846(.043)**	1.046(.099)	.793(.071)**
Identifying Program Problems	2,564	.824(.107)	1.022(.104)	.676(.071)**
Taking Corrective Action	2,568	.855(.091)	1.147(.129)	.760(.085)
Adopting New Approaches	2,546	.900(.097)	.867(.095)	.625(.069)**
Setting Performance Goals	2,517	.905(.124)	1.271(.157)+	.771(.078)**
Developing and Managing Contracts	1,846	.917(.169)	1.457(.327)+	.918(.187)
Streamlining Programs	2,363	.911(.116)	1.116(.196)	.823(.096)+

Notes:
Fixed Effects Proportional Odds Model; coefficients are odds ratios
Production bureaucracies are base case (=1)
** p<.01, * p<.05, + p<.10.
Robust standard errors in parentheses.

A broad examination of the odds ratio coefficients in table 5.4 reveals some striking patterns and provides evidence to address hypotheses H_{10a}-H_{10d}. For Production agencies, the

evidence shows Production agencies are about 37 and 47 percent more likely to adopt and use output measures than Craft and Coping agencies are, respectively, and are about 19 and 39 percent more likely to adopt and use outcome measures than Procedural and Coping agencies are, respectively. These results affirm the first two claims in hypothesis H_{10a}. Production agencies are more likely to use PI than Procedural and Coping agencies, but Craft agencies are even more likely than Production agencies to use PI except for developing program strategy or adopting new approaches to manage programs. Therefore, I reject the third claim in H_{10a} that Production agencies are more likely to use performance information than all other bureaucratic types are. That honor goes to Craft agencies.

For Procedural agencies, evidence in table 5.4 indicated Procedural agencies are about 26 percent more likely to adopt output measures than Craft agencies and about 35 percent more than Coping agencies. They are also about 28 percent less likely than Craft agencies and 19 percent less likely than Production agencies to adopt outcome measures. While the coefficient was not significant, the data suggest the possibility that Procedural agencies may even be somewhat more likely than Production agencies to adopt efficiency measures. Given the focus of Procedural agencies on processes and compliance with standards for conducting business, a natural by-product of this condition could be management interest in measuring the efficiency of internal agency processes. Procedural agencies are 10 to 20 percent less likely than Production agencies to make use of PI, but for eight out of nine PI uses, they are more likely to make use of PI than Coping agencies are. This evidence strongly supports hypothesis H_{10b}.

The evidence shows Craft agencies are 28 percent more likely than Procedural agencies and 48 percent more likely than Coping agencies to adopt outcome measures, are 37 percent less likely than Production agencies and about 25 percent less likely than Procedural agencies to

adopt output measures. They are far more likely to use PI in all categories of management decisions than Coping agencies are. The evidence therefore supports hypothesis H_{10c}.

For Coping agencies, the odds of adopting any performance measures decline from Production agencies about 30 to 50 percent across the board and are statistically significant for four out of five practices. Therefore, the evidence supports the claim in hypothesis H_{10d} that Coping agencies are less likely to adopt performance measures of any kind. For uses of PI, the significant results show Coping agencies are less likely than all other types of bureaucracies to use it for any kind of management purpose, and the results are statistically significant for seven of nine PI uses. The evidence strongly confirms hypothesis H_{10d}.

The results for Craft agencies for PI uses were contrary to expectations. Only two variables are statistically significant. They suggest informants based in Craft agencies are more likely than those in even Production agencies to report making use of PI for management decisions. The interviews with NASA PSMs (appendix R and chapter 4) reported that NASA has "…a robust system of program performance reviews and measures separate from GPRAMA, but it feeds into GPRAMA reporting." Note as well the large and significant coefficient of Craft agencies for "developing and managing contracts." This evidence suggests Craft agencies may have sedimented mechanisms for measuring the performance of their scientific programs and contract vehicles predating GPRAMA.

Based on these results, a claim can be made that Craft agencies are the most likely category of bureaucracy to institutionalize the use of performance information for management decisions, but that these uses are not necessarily linked to GPRAMA. If this is true, then a broader conclusion may be postulated: the likelihood an agency makes use of performance information in management decisions is partly a function of organizational identity. Agencies

whose identity is grounded in a mission that is primarily technical, scientific, and professional in nature are more likely to make use of performance information than agencies with general administrative or diffuse identities.

Chapter 6. Conclusions

Why Study Performance Management Institutionalization?

Over the course of a 36-year military and civilian career in the U.S. Government working for several agencies, I have observed the rise and fall of many management innovations. Cynics may refer to them as passing fads, but I am not able to so easily dismiss them. As a young Air Force captain serving in Japan in the 1990s, I was required to learn about and find ways to try to implement Total Quality Management (TQM) because, the thought went at the time, if TQM could make Japanese industry as successful and competitive as it was in those days, then it must be the key to unlocking improved performance in the U.S. Air Force. About a decade later, as a senior staff officer and assistant Director of Operations at a large Air Force center, I was placed in charge of my Directorate's internal controls program, a management control system designed to improve performance and fiscal accountability through definition and continuous monitoring of core mission processes, outputs, and outcomes. By that time, TQM was no longer in fashion in the Air Force. It had been replaced by a management control program focused more on accountability, compliance, and risk avoidance.

Several years after that, as a senior civilian in the Department of Defense, I observed the arrival of another impulse from the private sector to improve organizational performance, this time using the Six Sigma approach, and sent many members of my staff through training to become Green Belts and Black Belts. What I now find somewhat odd is that, until I attended the Army War College in 2010-2011 as a DIA civilian and took the elective on Defense Enterprise Management, I had never heard of GPRA, even though it had been around for 17 years! In the Department of Defense organizations in which I had served until then, if anyone was supporting GPRA, it must have been a small and specialized staff working in a small windowless office

deep within the Pentagon. GPRA-based performance management had not been institutionalized at the operational level of the department, as far as I was able to observe.

During each of these periods of my career, I was left with the sense that each wave of performance innovation had some merit and promised eventual performance benefits to the organization, but also that something about its implementation in the federal government eventually led to its early demise and eventual replacement with something else. What I did not recognize then, but with the benefit of hindsight and my education at Virginia Tech now can, is that each wave of innovation left behind something of value for the next, just as Paul Light observed about the tides of reform. As I see it, this is encouraging. It means that no reform effort is ever totally wasted. As long as some aspect of it becomes sedimented through the process of institutionalization, it will live on as an element of a future performance management regime in some manner. After completing this study, I have come to the conclusion this is the ultimate fate for GPRAMA as well as the government now moves into the era of "evidence-based" policies and practices.

But what about the rest of it, I wondered? The part that that did not survive the passage of time? What external governance and internal organizational conditions fostered its adoption and institutionalization in the first place? Could these same conditions in a similar configuration foster future reform? And, what changed over time to weaken manager and workforce support for the reforms? As a PhD student at Virginia Tech, I discovered I was not the only scholar who had asked exactly the same questions about the fate of different kinds of government reform, and how configurations of conditions in the governance and organizational environments interacted to sustain or terminate reforms. To fully understand the wide scope of these conditions and their possible effects on the institutionalization of reform, and to even understand what

institutionalization means in the context of public policy, I undertook an extensive literature review that helped me to comprehend the broad scope of the problem and apply some helpful theoretical frameworks from organizational theory and performance management scholarship for identifying the antecedents of performance management reform institutionalization and predicting their effects on reform outcomes. My scholarly aspiration for public policy and public management practice is that, by learning how and when to change these antecedent conditions through effective management and leadership strategies, one might well improve the prospects for the outcomes of public sector reforms, performance management or otherwise.

I discovered the literature on the institutionalization of performance management in the U.S. federal and state governments is extensive, but extremely diffuse. Chapter 2 presented a comprehensive overview and critical analysis of that literature, and observed most of it was conceptually "a mile deep and a foot wide," meaning most authors to date have engaged in deep theoretical or empirical research into the effects of individual antecedents like politics and resource dependency (P&R), oversight (O), goal clarity (G), a culture of empowerment and accountability (C), credible leadership commitment (L), employee training in performance measurement (T), capacity for evaluation of performance data (X), or the maturity of the measurement system (M). Some have published enlightening case studies into the mechanisms of and obstacles to institutionalization of either performance measures adoption (PM) or the use of performance information (PI) in a limited number of contexts. There have also been helpful meta-analyses of the growing public performance management literature, especially as it relates to scholarly interest in PI use for different management purposes.

Nevertheless, despite calls by many scholars for the Academy to develop better mid-range theoretical models and frameworks to explain the institutionalization of reform policies, I

did not find any theoretical frameworks or empirical research at the meso-level (cross-agency) in the U.S. Government that attempted to model the effects of all types of antecedents of performance management reform institutionalization at once, while controlling for common autoregressive conditions such as agency capacity for evaluation of performance, or measurement system maturity, that change over time as agencies sediment their performance management routines. I strongly suspected the literature's mostly piecemeal approach to separately analyzing the effects of individual antecedents resulted in systemic overestimation of the influence of the antecedents on GPRAMA institutionalization outcomes, because almost all of it overlooked conjunctural causation, causal asymmetry, and equifinality, all of which the literature, especially that of QCA methodologists, indicates are ubiquitous patterns of causality in social phenomena.

The risk to the validity of research findings from such overestimation is substantial. If everything is significant, then nothing is substantial. Indeed, initial regression probes I conducted during my PhD studies using individual factor variables and not a fully-specified multi-variable model estimated every single factor variable predictor, independently, was statistically significant. That is not very likely! So, I developed my own three-level inferential model that synthesized the primary threads of theory and empirical research using the Multiple Governance Framework (Hill and Hupe 2006), an extension of the Institutional Analysis and Development framework for public policy analysis, as a conceptual foundation, then overlayed the primary conditions described in the literature over the system, organization, and individual action situations at the directive governance action level. The analysis and findings in chapter 4 concerning the first research question were designed to make a contribution to the scholarship on public performance management by transcending the piecemeal strategies of previous

researchers and offer a more comprehensive analytic approach to estimate complex patterns of causation resulting from the concurrent interactions of multiple antecedents. My hope is that this contribution well be helpful to other scholars seeking to develop mid-range theories on reform policy institutionalization, and to practitioners who are trying to decide where to invest their limited time, attention and resources in order to advance performance management practices in their agencies.

My research into the history of the modern performance management movement in the U.S Government, in chapter 3, was designed as much to help me make sense of the often conflicting narratives in the literature about observers' perceptions of the effectiveness and social value of performance management systems as they were to help me ground the cross-sectional analysis of the contemporary contexts in 2017 and 2021 that followed in chapters 4 and 5. This research focused on historical evidence and scholarship from the GPRA, PART, and GPRAMA eras that revealed how indicators of P&R, O, G, C, L, T, X, and M, at different times and under different circumstances, affected the long arc of the institutionalization of performance management reform across the U.S. Government. I believe that my interviews with, and extensive archival records donated by, notable officials who were personally engaged in the institutionalization of GPRA, PART, and GPRAMA, and who now serve as fellows with NAPA, added both primary sources and immeasurable value and authenticity to the history section.

A second research goal in this dissertation was to investigate and empirically validate claims made by James Q. Wilson in his seminal treatise *Bureaucracy* (1989) concerning the effects of bureaucratic type on agency institutionalization of performance management. Chapter 2 showed that organizational identity is a growing field of inquiry within organizational theory, but enthusiasm for studying this phenomenon has yet to diffuse very far into the scholarship on

public management or performance management institutionalization. Yet, Wilson presented compelling theoretical predictions that bureaucratic identity should exert significant force on public manager decisions as to how and how extensively to adopt and implement performance measurement in their organizations, and on how performance information is used for decision making. Wilson's taxonomy is intuitively appealing, but what has been lacking in the literature is empirical evidence to back up Wilson's predictions about the effects of bureaucratic type on the institutionalization of performance management systems in public organizations. The findings for the second research question in this study represent an original empirical contribution to the scholarship to test Wilson's claims. They are intended, in part, to encourage fellow public management scholars to consider bureaucratic identity as an important antecedent of reform institutionalization, alongside external oversight, goal clarity, organizational culture, credible leadership commitment, employee training, organizational capacity for evaluation, measurement system maturity, and other conditions.

Summary of Research Findings

Conclusions for mixed methods studies should integrate the findings from the quantitative and qualitative research that was conducted. Table 4.1 in chapter 4 presents the relationship of the data sources and methods with elements of the research questions. A detailed discussion of the quantitative and qualitative research designs is presented in chapters 4 and 5. This section summarizes the findings, relates them to the study's hypotheses, and provides an integrative analysis.

Quantitative Findings

For research question one, tables 4.9 and 4.10 in chapter 4 presented the consolidated results of this study's quantitative modeling efforts. The study finds that a culture of

empowerment and accountability, the capacity for evaluation, and the maturity of the performance measurement system are all positively associated with the adoption of performance measures, when controlling for the effects of other variables postulated in the literature, in U.S. Government agencies. These results affirmed hypotheses H_{4a}, H_{7a}, and H_{8a}, respectively. I also found external oversight, goal clarity, and credible leadership commitment, all commonly reported in the literature as significant antecedents of performance management institutionalization, were not significantly associated with performance measure adoption in the U.S. Government when controlling for the influence of other organizational conditions, thereby disconfirming hypotheses H_{2a}, H_{3a}, and H_{5a}, respectively. Training in performance management was statistically significant for agency performance measure adoption, but analysis of the strength of the correlation and the absence of any significant association between it and the adoption of any single performance measure led me to dismiss it as insubstantial, disconfirming hypothesis H_{6a}. Yet, there is sufficient evidence, I believe, to not entirely rule out an association, particularly given the adolescent state of the research in this area.

For agency use of performance information (PI), I found that external oversight is not associated with the use of PI in agencies, supporting hypothesis H_{2b}. Agency goal clarity was not associated significantly with the use of performance information, disconfirming hypothesis H_{3b}. An agency culture of empowerment and accountability and the capacity for performance evaluation are positively associated with PI use across the federal government, strongly supporting hypotheses H_{4b} and H_{7b}. Credible leadership commitment was not significant for PI use generally, thereby disconfirming hypothesis H_{5b}, but the study finds it is significant for management decisions to use PI to develop program strategies, set program priorities, allocate resources, identify problems, take corrective actions, and set performance goals. Somewhat

unexpectedly, the evidence suggests performance management training is not associated with the use of performance information in agencies, disconfirming hypothesis H_{6b}. The evidence also indicated there was no significant association between the maturity of an agency's performance measures and its use of PI, disconfirming hypothesis H_{8b}.

Excursions from the base models 1, 2, and 3 that excluded the two autoregressive control variables of evaluation capacity (X) and measurement system maturity (M) exposed additional findings regarding what antecedents matter at different stages of institutionalization in agencies. The exclusion of X and M effectively simulated organizational environments in which these conditions were absent; specifically, agencies in the early phase of reform institutionalization, where X and M are not present. In the excursions for models 1 and 2, I detected a statistically significant relationship for goal clarity with both indices of GPRAMA institutionalization, but more so with the adoption of performance measures. The excursions for model 3 increased the number of performance management practices that are associated with goal clarity from 2 to 9, and 7 out of the 9 were significant. Therefore, I propose a possible explanation for these results is that agencies with adequate capacity for evaluation and mature measurement systems already incorporate a number of clear goals in their PMS, so once these are institutionalized, future reform outcomes depend less on future marginal improvements in goal clarity. Therefore, agencies which have less advanced infrastructure for evaluating performance, and have immature measurement systems, may advance GPRAMA institutionalization initially by focusing on clarifying goals for their organization and its programs first. Those which have sedimented clear goals in their measurement practices should then focus more effort to build a climate of accountability and empowerment in their cultures.

The results of the model excursions also indicated that cultural change is always important for performance management reform, but even more so when an agency has little evaluation capacity or immature performance measurement systems. Once these are institutionalized, these organizational conditions replace some of the need for cultural improvement. Without the presence of measures for evaluation capacity or measurement system maturity, credible leadership commitment in the model 3 excursion became significantly associated with ten GPRAMA practices. A possible interpretation of this result is that the contributions of credible leadership are most impactful in the early stages of reform institutionalization. In agencies with more fully institutionalized practices for evaluating and measuring performance, credible leadership commitment may be perceived by federal managers as a less salient condition for adopting new performance measures and using performance information. The study therefore finds that the commitment of agency leaders is a critical success factor during the habitualization phase of institutionalization, as Tolbert and Zucker (1996) proposed. The evidence here suggests agency senior leaders must demonstrate their commitment early on if they want to advance the institutionalization of reform to the objectification stage. Once an agency sediments performance management practices into daily measurement routines and makes the necessary investments in evaluation capacity, leadership commitment becomes less necessary. Once the change is made and deeply institutionalized, leaders can refocus their time and attention to other matters because the PMS is on auto-pilot. In other words, credible leaders "prime the pump" for performance management institutionalization in their agencies. Therefore, as noted in the literature review, this evidence supports the claim that public leaders can influence the institutionalization of performance management reform in their agencies, but their impact is strongest in the early stages of institutionalization. It may be

that transformational leaders have potentially greater impact than transactional leaders do during this phase, but that is a question for another research project. The literature on integrative leadership suggests this may not necessarily be true.

A quantitative approach also proved useful to demonstrate the variation of the effect of agency identity on PM and PI. Modeling approaches employed sought to measure random (cross-agency) effects on the two GPRAMA institutionalization indices. The results showed substantial variation between all 24 CFO Act agencies for both indices, but the effect was more than twice as strong on adoption of performance measures as it was for use of PI. Analysis of federal agencies clustered by Wilson bureaucratic type (tables 5.1 and 5.2 in chapter 5) showed that the type of bureaucracy an agency is matters for GPRAMA institutionalization. Mean random effect scores by Wilson bureaucratic type were all positive for Production, Procedural, and Craft agencies, but Coping agency effects were strongly negative, empirically validating Wilson's claims. The evidence confirmed hypotheses H_{9a} and H_{9b}.

I also wanted to know whether agency identity might affect agency adoption of specific kinds of performance measures and use of PI for specific management purposes. Results at appendix F and appendix G and the detailed analysis in chapter 5 provide mixed evidence in support of Wilson's prediction that the type of bureaucracy an agency is matters for its adoption of performance measures and how it uses PI. The evidence leads to a confident validation of Wilson's claims concerning Production and Coping agencies, but was less certain for Craft and Procedural agencies, likely due to insufficient evidence for these bureaucratic categories in the survey dataset to produce adequate levels of subset consistency in the fsQCA. It seems probable that a larger survey sample for these bureaucracy types would generate more consistent results and findings. Based on the available evidence, I conclude Production agencies are more likely

than all other types of bureaucracies to adopt all kinds of measures and, except for Craft agencies, to make greater use of performance information than other agency types, partially confirming hypothesis H_{10a}. Coping agencies are the least likely to adopt all types of measures or to use PI, confirming hypothesis H_{10d}. There was also some unexpected evidence that indicated Craft bureaucracies are the most likely type to make regular use of PI, partially confirming hypothesis H_{10c}.

In addition to analyzing the effects of individual agency identities, and as a cross-check for the findings from the cluster analysis above, I investigated what effects there might be for the Wilson bureaucratic types directly by using bureaucratic type as a nominal predictor variable. The results at table 5.3 in chapter 5 confirmed the cluster analysis results as well as Wilson's claims regarding the effects of different bureaucratic types on agency adoption of performance measures. The magnitude of the effect on performance measures adoption was statistically significant but small: the model indicated bureaucratic type, by itself, explained only about 1 percent (r^2=.0099) of the variation in PM, and its magnitude on use of PI was smaller still: bureaucratic type explained less than half of a percent of agency variation in the use of PI (r^2=.0028). Therefore, I conclude bureaucratic type has some influence on GPRAMA institutionalization outcomes, but it cannot explain them by itself. In a causal combination, its influence is easily overshadowed by other conditions. Nevertheless, the results did offer one unexpected and interesting finding: the evidence indicated the identities of Craft agencies were positively and significantly associated with the use of PI. This suggested some characteristic of Craft agencies makes them more likely to make use of PI than other bureaucratic types. To determine what this was, the qualitative research offered a new hypothesis, described in the qualitative findings below.

Finally, to support public management practice, I investigated the likelihood that different bureaucratic types would institutionalize the fourteen GPRAMA practices associated with PM and PI. The results at table 5.4 and analysis in chapter 5, with one exception, closely matched the patterns Wilson's taxonomy predicted and affirmed most claims in hypotheses $H_{10}a$-H_{10d}. The exception for $H_{10}a$ was, again, the surprising finding that Craft agencies are more likely to make use of PI than even Production agencies are, except for developing program strategy and adapting new approaches to manage programs. The relative paucity of quantitative and qualitative data for Procedural and Craft agencies, however, weakens confidence in these findings and points to the need for future research to validate the tentative results of this study.

Qualitative Findings

As noted in chapter 4, I pursued qualitative research in order to extend the quantitative research from the meso (cross-agency) level to the micro-level within federal agencies. My modest goal was to discover empirical examples of the conditions in the inferential model and to better understand how they interacted with one another, agency mechanisms for institutionalizing GPRAMA, and how the attitudes and biases of agency PSMs might mediate the institutionalization of GPRAMA practices. This investigation also extended the scope of studied institutionalization antecedents by adding politics and resource dependency (P&R) to external oversight as an additional system action situation condition for the inferential model described in chapter 2, as this condition could not be effectively operationalized using the large-N dataset used for the quantitative models. The qualitative research consisted of a 15-minute online survey taken by 20 performance system managers at 12 CFO Act agencies, and an approximately 1-hour online structured interview with six mid-level, senior, and senior executive performance system managers (PSMs) in five CFO Act agencies. The qualitative survey data was used in a fuzzy set

Qualitative Comparative Analysis (fsQCA) and blended with the results of the PSM interviews into five agency case studies (appendices Q through U and case narratives in chapter 4). Questions concerning PSM incentives, attitudes and biases were reserved for the interviews.

The desired outcome of the qualitative research was to discover causal configurations that would lead to positive and negative outcomes for agency institutionalization of GPRAMA. Using fsQCA methods to analyze the 2021 survey data, the study found that there were one necessary and three sufficient configurations of conditions (causal pathways) leading to the outcome of performance measure adoption that had levels of consistency and coverage above the study thresholds (model 1A in table 4.21 and table 4.18).

Solution 1, exemplified by the HUD, NASA, and Treasury cluster, was characterized by the absence of politics and resource dependency (~P&R) and the absence of a mature performance measurement system (~M). This configuration is not sufficient to bring about the outcome because its subset relationship with the set PM is one of necessity, which means additional conditions must explain performance measure adoption at HUD, NASA, and Treasury. But which conditions are hiding beneath the noise of the qualitative survey data? To answer this, we follow the evidentiary bread crumbs beginning with the kinds of performance measures agencies are likely to adopt and use. The estimates of the fixed effects of agency identity in appendix F showed all three agency identities had a strongly positive likelihood of adopting efficiency measures, and NASA and Treasury were among the most likely within their bureaucratic type categories to adopt quality and outcome measures (NASA) and customer service and quality measures (Treasury).

The results from the interviews with the PSMs in these agencies in appendices Q, R, and S, and the case narratives in chapter 4, shed light on possible conditions that could explain the

high likelihood for their adoption of these measure types. For HUD, adopting efficiency measures is only natural for a Procedural-type agency devoted to administering billions of dollars in funding via state and local grantees. This finding for HUD directly confirms the specific prediction of Jennings and Haist (2004) that Procedural agencies will tend to embrace efficiency measures. As already noted, it is theoretically challenging for Procedural agencies to observe and measure their outcomes, and HUD efforts to employ other kinds of measures are often impeded by the inability or unwillingness of external grantees to implement HUD measurement systems. For NASA, a Craft agency, program outcomes are directly observable and the quality of NASA R&D programs directly affects program outcomes. Even a small flaw in the quality of work that is overlooked can lead to disaster, as noted in Chapter 3 concerning the Challenger disaster in 1986. For Treasury, the PSM interviews detailed Treasury's strong focus on customer service and quality measures, both heavily influenced by the IRS and the department's public-facing services. Furthermore, the PSMs at Treasury emphasized the steady and vocal commitment of senior leaders to continuous improvement, which may explain why Treasury is more likely than the average agency to adopt efficiency measures. The overall pattern is clear: agencies tend to adopt measures which they can observe and which conform with their organizational identity. These results validate the claims of Spekle and Verbeeten (2014) that public organizations will tend to adopt performance measures that reflect the nature of the activities of the organization.

Three sufficient solution pathways leading to a positive outcome for PM were the presence of both P&R and C (DOC, DOJ, and USAID were ideal types), the presence of both P&R and X (DOD, HHS, and USAID were ideal types) and the presence of both C and X (DHS, SBA, and USAID were ideal types). The presence of USAID in all three configurations, as

shown in chapter 4, indicates some cases straddle different configurations, which is not unexpected given the nature of conjunctional causation. A reasonable interpretation of these findings by solution is that 1) a strong culture of accountability and empowerment may give agencies a means to assimilate and manage political influences and resource constraints, leading to a proliferation of performance measures that respond to requirements of external stakeholders, 2) agencies which have made extensive investments into evaluation capacity in the past and are exposed to high levels of P&R may not require as much C to be responsive to performance reporting requirements, and 3) in line with the quantitative research findings, combining C and X seems to be a strong recipe for success because it combines manager commitment to accountability and empowerment with organizational capacity to support performance measurement. Based on both the quantitative and qualitative research, C and X, when combined, create strong incentives for agency leaders to adopt a wide range of performance measures.

The findings from the qualitative research affirmed hypothesis H_{1a} on the effects of P&R, with an important qualification. The evidence from the 2021 survey, fsQCA models, and interviews with federal agency performance system managers (PSMs) showed that the absence of P&R is a condition in a necessary causal configuration for the adoption of performance measures (PM) in agencies with relatively less mature performance measurement systems (with HUD, NASA, and Treasury being ideal types for the causal configuration), but also that the presence of P&R is associated with the adoption of performance measures in agencies with a culture of accountability and empowerment (DOC, DOJ, and USAID are the ideal types) or substantial capacity for evaluating performance data (DOD, HHS, and USAID fit this profile well). This finding is similar in nature to the findings of the excursions of the fitted quantitative models insofar as it also makes a distinction between agencies based on their level of past

institutionalization of GPRAMA practices. This may be an important finding for the GAO and other oversight authorities. The relative presence or absence of P&R in federal agencies can affect their institutionalization of GPRAMA practices given certain conditions. The influence of P&R and mediating organizational conditions for GPRAMA institutionalization should be a consideration and potential qualification for oversight evaluations of agency implementation of Congressional and OMB reform mandates. All agencies are not equally situated to implement GPRAMA mandates.

To improve the institutionalization of performance measures, I conclude that federal managers should develop a strategy that is tailored to the unique conditions of their agency. If the agency experiences substantial levels of P&R, an appropriate strategy would be to work on making the agency's culture more accountable and empowered, or to invest more resources into capacity to evaluate performance data. Where P&R is less of an issue for the agency, which the PSM interviews indicated is the case most of the time, manager attention and agency resources should be focused on both C and X. Based on the QCA findings as well, these options appear to offer the greatest potential for return on time and effort, as a general recommendation.

For hypothesis $H_{1b,}$ the qualitative evidence generated by the fsQCA using the 2021 PSM survey dataset identified two sufficient causal configurations (table 4.23 in chapter 4) that result in federal manager use of performance information. In the first solution, the absence of P&R is a condition in a sufficient causal configuration with the absence of a culture of accountability and empowerment but with the presence of credible leadership commitment for use of PI for management decisions. In the second, the causal configuration was an absence of a culture of accountability and empowerment combined with the presence of credible leadership commitment and the presence of capacity for evaluation. Contrary to the results of the quantitative research,

based on the qualitative evidence, an organizational culture of accountability and empowerment does not appear to be a significant factor for manager use of performance information. Instead, demonstrated commitment by senior leaders and investments into evaluation capacity are the more important conditions. Per table 4.13 in chapter 4, for those federal agencies with a mean response to the 2021 survey that was above the cross-agency sample mean for the use of performance information, a majority of their agency mean responses for both L (4/7) and X (4/7) were also above the sample mean for those conditions. Expressed normatively, one may conclude that public manager commitment to performance management, coupled with investments into evaluation capacity, positively affects agency use of performance information in agencies. Furthermore, the recurring emphasis in the literature on senior leader efforts to change agency culture as a basis for improving manager use of performance information is, based on this evidence, likely overstated.

The fsQCA also modeled a single sufficient configuration for agency non-adoption of performance measures: the absence of P&R and the absence of capacity and the presence of a mature performance measurement system. The absence of politics and resource dependency, combined with the absence of capacity for evaluation, in agencies with mature performance measurement systems, may better explain why those agencies fail to adopt a wide range of performance measures. The finding concerning the presence of a mature performance measurement system (M) contradicts some empirical studies in the literature (e.g. Proeller, Kroll, and Meier 2012), but is explainable by the fact that the absence of M, coupled with the absence of P&R, was shown to be a necessary condition for the adoption of performance measures in the first solution. The model therefore suggests that the presence of M, and not its absence, makes adoption of new measures less necessary. This seems logical. Finally, the fsQCA modeling

identified no causal configurations having a consistent subset relationship with the set of agencies not using PI. Additional research is therefore suggested for this causal configuration.

Assumptions and Limitations

The quantitative research design made a number of assumptions and had important limitations which should be considered in assessing the validity of the study's findings and conclusions. Because the quantitative portion of the study used cross-sectional data from 2017, it cannot be easily generalized to other periods. This fact was one reason I chose to make this a mixed methods study that included current evidence. For the quantitative research, the data were collected by the GAO from federal managers between November 2016 and March 2017, shortly after a national election when there was substantial turnover of political appointees which may or may not have affected the perceptions and responses of Senior Executive Service and agency managers. A survey performed midway between Presidential elections might have produced substantially different results. The GAO dataset did not provide any individual data concerning respondents apart from their designation as senior executives or not, so typical control variables for age, sex, gender, race, education and other factors commonly modeled (Kroll 2015b) could not be included in the analysis. The use of two autoregressive factors, capacity for evaluation and performance measure maturity, was designed to model two controls based on their *a priori* existence and contribution to manager perceptions about their agency's performance management system. In the absence of data regarding demographic characteristics, however, it cannot be known for certain if these serve as sufficient controls, so quantitative results may be biased. Nevertheless, as cited, the same GAO management dataset from prior years has been used successfully by top scholars in the performance management field for cross-sectional

studies under the same limitations and threats to validity, and this study has applied aspects of their methodologies.

Next is the source and structure of the quantitative research dataset, a GAO survey of over 4,000 federal managers. The study assumes the survey administered by GAO is an agency-stratified random sample survey, based on GAO report GAO-17-775 (U.S. Government Accountability Office 2017a, 4, 96). The author has no means to independently verify this claim. The method GAO used to select recruits from an OPM dataset seems straightforwardly random (97). In some cases, some groups of managers in some agencies were not in the OPM dataset used by GAO; GAO reached out to those agencies to request their support to recruit those officers (98). GAO asked the Department of Justice to administer the survey internally for the GAO to protect agent privacy (98). It is certainly possible some self-selection effects may have occurred due to intentional agency management agendas or unrecognized personal biases. The GAO undertook several measures to assure the accuracy, completeness and reliability of the data (98-99). Because the GAO survey has been used in research articles approved for publication in top public administration journals as noted in the literature review, it seems to be a sufficiently reliable and valid source of quantitative data for the purposes of this study as well.

The structure and content of the GAO dataset, nevertheless, limited is usefulness in important ways. First, the data is uniformly stratified by CFO Act department or independent agency (24 in all). This means it is not possible to sub-stratify the data to the sub-agency level, for instance at the level of the IRS within the Department of Treasury or of FEMA within DHS. As a result, the study perforce compares enormous organizations like the Department of Defense and Department of Homeland Security with very small organizations like the National Science Foundation. It is clear that within the large and aggregated departments, one may find elements

which come closer to being ideal-type Wilson bureaucracies; therefore, the blending of a variety of different elements within a department made the results and findings for larger departments more ambiguous and probably less likely to register in the fsQCA as ideal types for different causal configurations. Adding a sub-departmental organizational designation field to the GAO dataset would allow future researchers to compare departmental sub-organizations with better defined, more uniform, and contrastable organizational identities. Second, the dataset does not include some commonly collected workforce information on survey respondent characteristics such as age, gender, racial identification, professional/occupational category, etc. This study was thus unable to control for these conditions in the quantitative models. This is significant because any of these conditions may affect manager responses to accountability-based performance management reforms, or to survey questions.

Furthermore, any self-reported data from a survey has some inherent limitation that should be noted. As Destler (2016) observed, ". . . self reports capture espoused values, but may not reflect deeper, tacit assumptions and the realities of organizational practice" (203). Also, a potential limitation of any study based on cross-sectional survey data is the possibility of common source bias (CSB) when the data for the dependent and independent variables come from the same source, which can lead to spurious findings (Meier and O'Toole 2013, 431). How survey questions are phrased matters. Questions about respondent perceptions on their reactions to strategy, support from the environment, manager networking or observable behaviors were found to be much less prone to CSB than questions that require the respondent to assess quality, environmental buffering strategies, diversity, or environmental manipulation (Meier and O'Toole 2013, 443). Fortunately, in the GAO survey, all of the questions fall into the former categories, not the latter. The GAO questions are based on facts and manager perceptions about agency

work practices, neither of which carry much in the way of implications for organizational or personal reputation or career advantage for a positive response. If anything, the negative bias in the model 1 and 2 residuals described in the research design in chapter 4 points to a probable deflationary scoring bias by federal managers.

In addition to these considerations about the risks to validity due to CSB effects, the general concern of methodologists in public administration is with the inflation of survey responses to questions about subjective participant assessments of organizational or individual performance (George and Pandey 2017, 247). This makes sense if a respondent's performance as an employee or as an organizational leader might be questioned based on the survey findings. They may have a personal interest in the study's outcome. As no GAO survey question asked participants to rate performance, and the survey was administered anonymously, this is not a concern here. Moynihan and Kroll (2016) dismissed the risks of CSB for their analysis of earlier GAO manager surveys with the view that there are no prospective payoffs for federal managers to inflate their responses to the anonymous survey. Their argument applies equally well here. Dependent variables in this model are not dimensions of performance per se, but manager perceptions about what they have observed in their organizations. The questions offer no veiled incentives for respondents to inflate their anonymous survey scores. The absence of such incentives should help to mitigate any worries about the presence of CSB in survey data (George and Pandey 2017, 254). Additional reassurance comes from GAO statements that survey questions were pre-reviewed by "…subject matter experts, a survey specialist, and a research methodologist" and adjusted after received feedback from pre-testers (Government Accountability Office 2017c).

Another assumption concerns the validity of the latent variables and the data in them. The latent variables, which drew observed survey question response data from the GAO dataset, underwent transformations to make them less abnormal, then standardization to make them comparable and more interpretable, and then multiple imputation to fill in missing data to increase sample size. The study assumes these statistical techniques to improve validity and interpretability, all recommended by professional statisticians and scholars of public administration, did not also do violence to the underlying observed GAO data. While the various tests for violations of OLS assumptions demonstrated sufficient results to proceed with fitting models using robust standard errors, a non-statistician public manager such as the author may be left with a sense of foreboding that somehow these techniques may have corrupted the data in some way. My assumption, and sincere hope, is that they did not.

While the study models employ almost as many predictor variables as the primary threads of debate about reform institutionalization in order to put the major claims of the performance management literature to the test, it is conceptually simple and analytically sufficient for the goal of this correlational phase of the dissertation to isolate the prime predictors to use in the qualitative phase. In order to isolate the effects of each predictor on each dependent variable, the models omit interaction effects that could more realistically describe, for example, the mediating effects of organizational factors on the relationship of oversight with the adoption of performance measures, or the indirect but valuable contribution of performance management training to increasing empowerment in agency culture or to improving goal clarity. All of these and more would be worthy research endeavors in their own right that could be undertaken within the inferential framework and would represent an enhancement to the simple, but multiple, bivariate correlational findings in this study. These quantitative models are therefore a

prototypical, limited, yet fairly comprehensive, analytical framework of potential value for scholars who wish to further investigate the multiple mediating relationships between the system and organization action situation conditions on performance management or potentially other forms of government reform institutionalization.

On their own, the quantitative findings cannot prove causality, merely highlight which system and organizational action situation conditions were most correlated with GPRAMA institutionalization indicators based on perceptual and cross-sectional data of federal managers in 2017, while controlling for the other conditions. It is the statistical significance and general magnitude of the covariance of the manager perceptions about these conditions within agencies and between agencies and their operating environment, the system action situation, which is of greatest theoretical interest as well as instrumental interest for selecting significant system action and organization action situation conditions for further investigation in the qualitative portion of the research.

Like the quantitative work, the qualitative research in this mixed methods study made several important assumptions, and limitations must be acknowledged. The study was limited by finite access to PSMs at federal agencies during a pandemic and the start of a new presidential administration in the spring of 2021, and by respondents' willingness to freely share their perceptions regarding the institutional and organizational conditions at their agency and their effects on their decision making for implementing GPRAMA. In particular, the size of the agency PSM interview cohort, six employees, was much smaller than planned, so data collected from PSMs was adequate to provide a degree of empirical validation and triangulation for the 2021 survey and fsQCA results, but it was not sufficient in its own right to permit generalization about the effects of inferential model conditions or the behavior of bureaucratic types across the

government. In addition, there are no state or local government respondents in this study, so findings may only be generalizable to similarly-situated national governments without additional empirical research using the inferential model at those levels.

The fsQCA method, while offering a systemic and replicable approach to structure comparative case analyses, also has certain inherent limitations that imply any effort to generalize findings must remain "modest" (Berg-Schlosser et al. 2009, 17). Effective use of fsQCA as a methodology to analyze causal patterns depends on the researcher making several assumptions and presenting them in a transparent manner. Foremost among these are defining thresholds for calibrating levels of membership of cases in sets, as this study did in appendix P. Calibrating thresholds for the very observable boiling and freezing points of water is straightforward, but it is much more difficult to propose and justify suitable thresholds for complex and often hard to observe social phenomena, which may explain the rarity of calibration in social science research (Ragin 2008, 72). QCA methodology recommends researchers invoke theoretically or empirically-informed references for calibration (e.g. Schneider and Wagemann 2013; Greckhamer et al. 2018, 488). Because no such references or benchmarks exist for setting thresholds for full and partial membership in the set of agencies which have fully institutionalized GPRAMA practices or which have an accountable and empowered cultural climate, for instance, I had to adapt methods used by acclaimed QCA scholars in the same predicament who have published successful peer-reviewed journal articles that combined case clustering and logical breakpoints with theoretically-developed rules as criteria for calibration. As such, all thresholds in the QCA are based on theoretical assumptions and empirical patterns and subject to review, criticism, and improvement. As theory on set membership criteria develop in the future, it is nearly a certainty the thresholds set in this study will be adjusted. The

method used to set thresholds in this study could serve as a starting point for that discussion. How can we know an agency has a culture that is fully accountable and empowered, or that it has sufficient capacity to evaluate its performance? How much is enough and what are the units of measurement?

The qualitative data sources were federal PSMs who, under no compulsion to participate and with no particular incentive to do so, volunteered to take an online survey. Six out of ten surveyed PSMs responded positively to invitations to interview after they had completed the surveys. There is every reason to believe participants who were intrinsically motivated to support performance management self-selected into both the survey and interviews, with those interviewed being in all likelihood among the most intrinsically motivated PSMs in the government. The author found many of the study volunteers to be strong believers in good government philosophy, which includes a high degree of transparency for the citizenry. While this study was not about public service motivation (PSM) per se, based on what I experienced interacting with federal officials, future quasi-experimental research might fruitfully test the degree to which PSM affects public manager response rates to academic surveys. From my anecdotal experience, I have no doubt it did so in this study, and positively. Self-selection by performance management advocates and optimists in the survey and interviews was an unavoidable reality in this research. With that in mind, survey and interview findings should be considered as generally inflated, and thus a limitation of the qualitative research.

The Way Ahead for Performance Management in the U.S. Government

If this study had been conducted in 1995, 2000, or 2005, the research findings on GPRA institutionalization would have been preliminary. Agencies were continuing to learn and trying to advance their performance management reform practices through the 3-step process of

habitualization, objectification, and sedimentation (Tolbert and Zucker 1996), as described in chapters 2 and 3. It is now the fall of 2021, the federal government has recently begun to expand reform efforts beyond the traditional conception of performance management, with its core values of agency and program accountability, transparency, and target-setting, to better integrate enterprise risk management, learning agendas, and evidence-based practices (Kamensky 2021, 38-42). Improved management of data has emerged as an important reform value and mandate, as reflected in the Foundations for Evidence-Based Policymaking Act of 2018 (Pub. L. 115-435). In the Biden Administration, OMB is seeking to achieve improved program outcomes through the institutionalization of the Evidence Act (Young 2021), even as GPRAMA remains far from fully sedimented across the government, as this study has shown. This may be interpreted as a sign that federal leaders and managers are now laying a new path for managing performance that replaces the paradigm of accountability- and compliance-oriented performance improvement with a new one that employs science and evidence-based decision making to improve program outcomes. In light of the history of government reforms, it is probable that GPRAMA itself will soon become merely the newest layer of sedimented performance-oriented reforms in the federal government.

The timing for the increased stakeholder attention and interest in these new practices over the past five years, during which GAO reports revealed stagnating agency progress in the implementation of GPRAMA and agency use of performance information for decision making, cannot plausibly be fortuitous. The rise of these new reform initiatives is an indication that GPRAMA-specific practices have achieved their maximum realistic level of sedimentation (still far short of full institutionalization at all levels), and that the next "tide" of reform has begun to rise, in part, due to the creeping realization that federal agencies have gone as far as they are

willing or able to in order to institutionalize GPRAMA practices. Like all previous tides, the new one builds on the institutionalized practices of its predecessor. At such an ideal juncture between two eras of reform, I perceived a timely opportunity to identify which system and organization action conditions were most influential during the endgame for GPRAMA institutionalization. Those same conditions will continue to influence future reforms, and the institutionalization of the evidence-based reform tide would benefit by the knowledge of the forces these conditions exerted on the last tide.

Recommendations for Future Research

I intentionally chose what I thought were challenging research questions for this dissertation in order to offer several theoretical, methodological, and empirical contributions to scholarship. Through my review of the literature, I perceived several lacunas on the institutionalization of performance management reform, and aimed at using this dissertation as a starting point for others to frame and pursue future research on reform policy institutionalization matters. First, as chapter 2 showed, there is a recognized need for better integration between organizational theory, especially the branches concerning institutionalization and organizational identity, and public management scholarship. Studying the institutionalization of GPRAMA provided a useful and interesting empirical context to study both of these vectors of organizational theory and apply them during research into a public management problem of importance. I certainly make no claim that the methods I used to do this were in any sense optimal, but they at least offer an approach that is transparent and replicable for fellow scholars who may wish to pursue additional micro-level research to discover intra-agency mechanisms and protocols that advance reform policy institutionalization from habitualization to objectification to sedimentation.

Second, again through my review of the literature, I found a wealth of theoretical and empirical studies focused on a single or at most two conditions that predict GPRAMA institutionalization, at either the system or organization action situation levels, but none with the insane ambition, borne of a wish to develop mid-range theories about reform policy institutionalization to support public management scholarship, to try to consolidate all major lines of theory regarding performance reform institutionalization into a single inferential model that controlled for the effects of all of these conditions simultaneously. In addition, I did not find a single study that accounted for the effects of conjunctural causation, causal asymmetry, and equifinality on the institutionalization of performance management reform, but learned why this is probably the case: public management scholars have yet to fully embrace QCA methods, even though they are designed to do precisely these things. Based on this review, I believe this study presents an original contribution to all fields of scholarship that use QCA techniques, as well as to the public management and governance scholarship. Future research may wish to validate my operationalization of the constructs in the inferential model (both the variables I used in the quantitative models and the conditions in the fuzzy-set QCA), replicate and compare the present study using future GAO or other datasets, or focus on interactions between two or more variables or conditions on institutionalization outcomes. The least developed portion of the inferential model in this study, namely the mediating effects of federal manager attitudes and biases, could well be the most important condition. Obtaining some individual level data from federal managers through interviews proved to be possible, but it was quite challenging in terms of recruiting participants due to a wide range of factors beyond my control during the period of research. Future scholars may wish to explore the impact of these conditions using ethnographic methods to acquire valid data and apply more sophisticated psychological and psychometric

models to publish research findings in the rapidly expanding field of behavioral public administration.

Third, I was captivated by the taxonomy of bureaucratic types first proposed by James Q. Wilson (1989) in his canonical work of public management, *Bureaucracy*. I found several allusions to it in the literature on public sector performance management, and some claims that the "one size fits all" standard of federal performance management reform neglected to recognize the impact of bureaucratic type on agency implementation of GPRA and GPRAMA (e.g. Gueorguieva et al. 2009; Jennings and Haist 2004; Radin 2006). What was missing was a method to classify actual agencies based on current evidence and any systematic test of Wilson's taxonomy to show whether his claims regarding agency performance management behaviors held empirically. This study devised a method to classify federal agencies, using Wilson's insight that bureaucracies will tend to measure whatever they are able to observe, based on federal manager perceptions about the extent to which their agencies observe their own outputs and outcomes. The results of this part of the study broadly supported Wilson's claims, with some important caveats that may be unique in the case of the U.S. Government, such as the unexpectedly strong finding based on qualitative data concerning the use of performance information by Craft-type agencies. The amount of evidence available in this study permitted a high degree of confidence in the findings regarding both Coping and Production-type agencies, but a shortfall in evidence concerning Craft and Procedural agencies led to somewhat unreliable findings concerning those types. Future research could acquire additional evidence to improve the reliability of findings for those types and validate the findings regarding Production and Coping agencies, or explore alternative methods for classifying agencies using empirical data.

Final Thoughts: Lessons on Public Policy Reform for Government Leaders

To conclude this study, I would like to offer several insights I have gleaned through this research about the governance challenges of institutionalizing public policy reforms in the American constitutional republic. As the author of GPRA and GPRAMA legislation, Congress should come first. The history of GPRA, PART, and GPRAMA in chapter 3 revealed the limits to institutionalizing liberation management-style reforms posed by prior legislation designed to ensure fiscal integrity and equitable and effective human resources management. Federal managers at all levels have far less authority to reassign or dismiss personnel, or to reallocate resources, than their private sector counterparts do, the precepts of NPM notwithstanding. The failure of the Executive Branch to undertake GPRA management flexibility pilot programs in the 1990s bears stark testimony to this characteristic of federal agencies and how Congress oversees them. In this respect, repeated Congressional complaints about some federal agencies' lack of progress in instituting needed management reforms to improve program performance seems uninformed at best or disingenuous at worst.

Federal managers who lack sufficient authority and control over resources, including staffing levels, cannot be expected to act like private sector managers who have such authority and control to generate higher levels of organizational or program performance. Even if they have substantial discretion, the scholarship on policy implementation suggests they may have reason to use that discretion to impede or at least not advocate for reform policy implementation when they perceive it to lack meaning for society, their clients, or themselves (Tummers, Steijn, and Bekkers 2012), perceive potential role conflicts, threats to professional norms, conflicting social commitments, or ethical conflicts (Thomann 2015), or unacceptable professional or personal risk (Behn 2001). As a general precept, federal bureaucracies are best served by

employing integrative leaders (Moynihan and Ingraham 2004). But during the early phases of reform institutionalization, the charismatic traits of transformational leaders may be well-matched to the institutional conditions that require a more dynamic and engaged approach to launch the reform in the agency and inspire credible commitment to the reform from all managers.

This study has focused attention on the effects of senior leader credible commitment to performance management within federal agencies. Evidence from the history of performance management reforms in chapter 3 and the evidence concerning the absence of influence of external oversight on agency adoption of performance measures and use of performance information suggest Congress and oversight authorities may suffer from federal manager perceptions that they lack credible commitment themselves. Reasons why federal managers perceive this were exposed in this study. These include, for example, historical disinterest by Congress in using performance information reported by federal agencies per GPRA and GPRAMA requirements in the budget process (one of the core justifications for passing GPRA and GPRAMA in the first place), lack of response to findings in GAO reports on agency non-implementation of recommendations, and poorly-hidden individual legislator agendas to exploit performance management reform for political gain for themselves or their parties. Students of public policy may easily conclude that, in the final analysis, the real purpose of performance management reform in the American federal government is either symbolic, a signal from elected officials to constituents that they care about good government and how it spends voters' tax dollars, or instrumental, a useful strategy to help legislators get re-elected.

The Office of Management and Budget has played an important role since the first days of GPRA, and its influence on GPRAMA institutionalization continues to be felt in every CFO

Act agency. Federal PSMs interviewed for this study uniformly acknowledged that OMB's influence over their agency's decisions concerning GPRAMA implementation exceeded that of any other external organization or stakeholder. The history of PART revealed that OMB can play a more directive role over agency performance management systems using its budgetary authorities, but this role creates high risks of politicization for program decisions, and the literature found evidence this occurred during the Bush administration. Supporters of performance budgeting should take note and remember that budgeting is an inherently political process (Wildavsky and Caiden 2004). To protect its reputation and legitimacy, OMB political appointees should delegate as much authority as possible to career federal performance system managers and establish internal agency business rules that minimize the potential for political manipulation of GPRAMA-related processes.

Concerning federal agencies, the historical record in chapter 3, the 2017 GAO survey results, the 2021 PSM survey, and the interviews with agency PSMs showed they have significant autonomy to make decisions about how they institutionalize GPRAMA, so long as minimum legal compliance is achieved. Some agency PSMs even complained during their interviews that OMB gave certain agencies, DoD in particular, a "pass" on meeting certain minimum GPRAMA reporting requirements. This concern is actually encouraging, because it reveals an emerging interagency consensus to support GPRAMA below the OMB level and the commitment to hold other members of the performance management community accountable for the law's implementation. Very recent experiences of the federal performance management community with sudden and unexpected OMB reversals on GPRAMA administrative guidance at the end of the Trump administration demonstrated the power and sustainability of performance management practice institutionalization: many agencies pushed back on the OMB decision to

discard its longstanding administrative guidance to agencies and successfully got them reinstated early in the Biden administration. Simple path dependency cannot explain such bureaucratic advocacy. Many agencies have embedded performance management practices as a routine and have come to appreciate the public value of measuring program and agency performance. During interviews, many PSMs stated OMB guidance was an indispensable tool for them to motivate agency leaders and managers to respond to their GPRAMA-related requests and taskings. Once performance management practices become sedimented in an agency, eliminating them becomes highly unlikely, exactly as institutional theory predicts.

It is now time to come to the bottom line: practice recommendations for federal agency leaders and managers. This study revealed federal agencies have followed a finite number of pathways to performance measures adoption and performance information use. Leaders in agencies with performance management practices still at the habitualization stage of institutionalization can most effectively foster future use of performance measures by establishing the credibility of their personal commitment to performance management practices and improving the attitudes of their managers about the value of performance measurement. In other words, a consensus among agency leaders and managers to support policy reform is a necessary, but not sufficient, pre-condition for reform institutionalization. In addition, the quantitative analysis indicated that improving goal clarity was significantly associated with performance measure adoption in agencies with less-developed performance management systems. Leaders of agencies in the objectification stage of institutionalization can advance performance measures adoption by investing in additional capacity for performance evaluation and cultivate a sense of accountability and empowerment in their agency's culture, as long as delegated authority does not come at the expense of accountability.

To foster use of performance information by federal managers, agency leaders should first assess the level of politics and resource dependency (P&R) under which their agency operates. This is a system action condition whose presence or absence was identified in chapter 4 as a critical differentiator between causal configurations. Most agencies studied in the qualitative analysis reported little or no P&R, but for those that reported some, other conditions needed to exist for effective institutionalization. For agencies subject to little or no P&R, agency leaders have an opportunity to foster use of performance information by making their commitment to it credible. In addition to leader commitment, the quantitative findings suggested that improving goal clarity, offering performance management training, and working on transforming agency culture could be worth considering as complementary strategies during the habitualization phase. In contrast with the quantitative findings, the qualitative analysis concluded that federal leaders can foster greater performance information use primarily by investing in capacity for performance evaluation and strengthening credible leader commitment. In other words, achieving long-term change in agency cultures to be more results-oriented may contribute to the use of performance information over the long run, but it is not necessary to achieve near-term improvements. This should be an encouragement to federal leaders who are wary of the complexities and long timelines it typically takes to transform agency cultures. The ultimate goal for federal leaders should be to achieve full sedimentation of performance management practices that allows leaders to focus their attention on other priorities and their agencies not only to comply with legislative requirements, but also to use performance information as a helpful tool for more effective public management.

References

Amenta, Edwin, and Jane D. Poulsen. 1994. "Where to Begin: A Survey of Five Approaches for Selecting Independent Variables for Qualitative Comparative Analysis." *Sociological Methods and Research* 23 (1): 22–53. https://doi.org/10.1177/0049124194023001002.

Ames, Fred. 2015. "The Drive to Improve Performance in the Federal Government: A Longitudinal Case Study of Managing for Results." Dissertation. George Mason University. https://pdfs.semanticscholar.org/a986/c925197e7b1605ae53d6c7620c2590688b98.pdf.

Andersen, Simon C., and Ulrik Hvidman. 2021. "Can Reminders and Incentives Improve Implementation Within Government? Evidence from a Field Experiment." *Journal of Public Administration Research and Theory* 31 (1): 234–49. https://doi.org/10.1093/jopart/muaa022.

Andersen, Simon C., and Mads Leth Jakobsen. 2018. "Political Pressure, Conformity Pressure, and Performance Information as Drivers of Public Sector Innovation Adoption." *International Public Management Journal* 21 (2): 213–42. https://doi.org/10.1080/10967494.2018.1425227.

Anderson, Derrick M., and Justin M. Stritch. 2016. "Goal Clarity, Task Significance, and Performance: Evidence from a Laboratory Experiment." *Journal of Public Administration Research and Theory* 26 (2): 211–25. https://doi.org/10.1093/jopart/muv019.

Andrews, Rhys, George Boyne, and Richard M. Walker. 2011. "The Impact of Management on Administrative and Survey Measures of Organizational Performance." *Public Management Review* 13 (2): 227–55. https://doi.org/10.1080/14719037.2010.532968.

Andrews, Rhys, and Ahmed Mohammed Sayed Mostafa. 2019. "Organizational Goal Ambiguity and Senior Public Managers' Engagement: Does Organizational Social Capital Make a Difference?" *International Review of Administrative Sciences* 85 (2): 377–95. https://doi.org/10.1177/0020852317701824.

Andrews, Rhys, and Steven Van de Walle. 2013. "New Public Management and Citizens' Perceptions of Local Service Efficiency, Responsiveness, Equity and Effectiveness." *Public Management Review* 15 (5): 762–83. https://doi.org/10.1080/14719037.2012.725757.

Ansari, Shahzad M., Peer C. Fiss, and Edward J. Zajac. 2010. "Made to Fit: How Practices Vary as They Diffuse." *Academy of Management Review* 35 (1): 67–92. https://www.jstor.org/stable/27760041.

Argyris, Chris. 1991. "Teaching Smart People How to Learn." *Harvard Business Review* 69 (3): 99–109.

Arnaboldi, Michela, Irvine Lapsley, and Ileana Steccolini. 2015. "Performance Management in the Public Sector: The Ultimate Challenge." *Financial Accountability & Management* 31 (1): 1–22. https://doi.org/10.1111/faam.12049.

Ashworth, Rachel, George Boyne, and Rick Delbridge. 2007. "Escape from the Iron Cage? Organizational Change and Isomorphic Pressures in the Public Sector." *Journal of Public Administration Research and Theory* 19 (1): 165–87. https://doi.org/10.1093/jopart/mum038.

Ashworth, Rachel, Ewan Ferlie, Gerhard Hammerschmid, M. Jae Moon, and Trish Reay. 2013. "Theorizing Contemporary Public Management: International and Comparative

Perspectives: Theorizing Contemporary Public Management." *British Journal of Management* 24 (September): S1–17. https://doi.org/10.1111/1467-8551.12035.
Baldwin, Elizabeth, Tingjia Chen, and Daniel Cole. 2019. "Institutional Analysis for New Public Governance Scholars." *Public Management Review* 21 (6): 890–917. https://doi.org/10.1080/14719037.2018.1538427.
Barley, Stephen R., and Pamela S. Tolbert. 1997. "Institutionalization and Structuration: Studying the Links between Action and Institution." *Organization Studies* 18 (1): 93–117. https://doi.org/10.1177/017084069701800106.
Battaglio, R. Paul, Paolo Belardinelli, Nicola Bellé, and Paola Cantarelli. 2019. "Behavioral Public Administration Ad Fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations." *Public Administration Review* 79 (3): 304–20. https://doi.org/10.1111/puar.12994.
Beeri, Itai, Anna Uster, and Eran Vigoda-Gadot. 2019. "Does Performance Management Relate to Good Governance? A Study of Its Relationship with Citizens' Satisfaction with and Trust in Israeli Local Government." *Public Performance & Management Review* 42 (2): 241–79. https://doi.org/10.1080/15309576.2018.1436074.
Behn, Robert D. 2001. *Rethinking Democratic Accountability*. Washington, DC: Brookings Institution Press.
Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones. 2019. "Fixed and Random Effects Models: Making an Informed Choice." *Quality & Quantity* 53 (2): 1051–74. https://doi.org/10.1007/s11135-018-0802-x.
Bell, Andrew, and Kelvyn Jones. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3 (1): 133–53. https://doi.org/10.1017/psrm.2014.7.
Belle, Nicola, Paola Cantarelli, and Paolo Belardinelli. 2018. "Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions." *Public Administration Review* 78 (6): 828–40. https://doi.org/10.1111/puar.12960.
Benjamin, Lehn M., and Paul L. Posner. 2018. "Tax Expenditures and Accountability: The Case of the Ambivalent Principals." *Journal of Public Administration Research and Theory* 28 (4): 569–82. https://doi.org/10.1093/jopart/muy040.
Berg-Schlosser, Dirk, and Gisele De Meur. 2009. "Comparative Research Design." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, 19–32. Applied Social Research Methods Series 51. Thousand Oaks, CA: Sage Publications, Inc.
Berg-Schlosser, Dirk, Gisele De Meur, Benoit Rihoux, and Charles C. Ragin. 2009. "Qualitative Comparative Analysis (QCA) as an Approach." In *Configurational Comparative Methods: Qualitative Comparative Analysis and Related Techniques*, 1–18. Applied Social Research Methods Series 51. Thousand Oaks, CA: Sage Publications, Inc.
Bhanot, Syon P., and Elizabeth Linos. 2020. "Behavioral Public Administration: Past, Present, and Future." *Public Administration Review* 80 (1): 168–71. https://doi.org/10.1111/puar.13129.
Bingham, Andrea J., Sean Dean, and Jessica Castillo. 2019. "Qualitative Comparative Analysis in Educational Policy Research: Procedures, Processes, and Possibilities." *Methodological Innovations* 12 (2): 1–13. https://doi.org/10.1177/2059799119840982.

Bititci, Umit S., Patrizia Garengo, Aylin Ates, and Sai S. Nudurupati. 2015. “Value of Maturity Models in Performance Measurement.” *International Journal of Production Research* 53 (10): 3062–85. https://doi.org/10.1080/00207543.2014.970709.

Blatter, Joachim. 2017. “Truth Seeking AND Sense Making: Towards Configurational Designs of Qualitative Methods.” *Qualitative & Multi-Method Research* 15 (2): 2–14. https://doi.org/10.2139/ssrn.3450025.

Bouckaert, Geert, and John Halligan. 2008. *Managing Performance: International Comparisons*. New York, NY: Routledge.

Bourdeaux, Carolyn, and Grace Chikoto. 2008. “Legislative Influences on Performance Management Reform.” *Public Administration Review* 68 (2): 253–65. https://doi.org/10.1111/j.1540-6210.2007.00860.x.

Box, Richard C., Gary S. Marshall, B.J. Reed, and Christine M. Reed. 2001. “New Public Management and Substantive Democracy.” *Public Administration Review* 61 (5): 608–19. https://doi.org/10.1111/0033-3352.00131.

Boyne, George A. 2010. “Performance Management: Does It Work?” In *Public Management and Performance: Research Directions*, edited by Richard M. Walker, George A. Boyne, and Gene A. Brewer, 207–26. New York, NY: Cambridge University Press.

Braun, Dietmar, and Fabrizio Gilardi. 2006. “Taking ‘Galton’s Problem’ Seriously: Towards a Theory of Policy Diffusion.” *Journal of Theoretical Politics* 18 (3): 298–322. https://doi.org/10.1177/0951629806064351.

Breul, Jonathan D. 1994. “How ‘The Government Performance and Results Act’ Borrows from the Experience of OECD Countries.” In *Fulbright Symposium on Public Sector Reform*. Australasian Evaluation Society Collection. Brisbane, Australia.

Breul, Jonathan D., and John M. Kamensky. 2008. “Federal Government Reform: Lessons from Clinton’s ‘Reinventing Government’ and Bush’s ‘Management Agenda’ Initiatives.” *Public Administration Review* 68 (6): 1009–26. https://doi.org/10.1111/j.1540-6210.2008.00950.x.

Brudney, Jeffrey L., F. Ted Hebert, and Deil S. Wright. 1999. “Reinventing Government in the American States: Measuring and Explaining Administrative Reform.” *Public Administration Review* 59 (1): 19. https://doi.org/10.2307/977476.

Bryson, John M. 2011. *Strategic Planning for Public and Nonprofit Organizations: A Guide to Strengthening and Sustaining Organizational Achievement*. 4th ed. San Francisco, CA: Jossey-Bass.

Busuioc, Madalina, and Martin Lodge. 2016. “The Reputational Basis of Public Accountability.” *Governance: An International Journal of Policy, Administration, and Institutions* 29 (2): 247–63.

———. 2017. “Reputation and Accountability Relationships: Managing Accountability Expectations through Reputation.” *Public Administration Review* 77 (1): 91–100. https://doi.org/10.1111/puar.12612.

Cameron, A. Colin, and Pravin K. Trivedi. 2010. *Microeconomics Using Stata:* Revised edition. College Station, TX: Stata Press.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. 2. print; Reprinted from "Handbook of Research on Teaching". Boston: Houghton Mifflin. https://www.sfu.ca/~palys/Campbell&Stanley-1959-Exptl&QuasiExptlDesignsForResearch.pdf.

Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy*. Princeton, NJ: Princeton University Press.
Carpenter, Daniel P., and George A. Krause. 2012. "Reputation and Public Administration." *Public Administration Review* 72 (1): 26–32. https://doi.org/10.1111/j.1540-6210.2011.02506.x.
Carrera, Leandro N., and Marina Angelaki. 2020. "The Diversity and Causality of Pension Reform Pathways: A Fuzzy-Set Qualitative Comparative Analysis." *Journal of Social Policy* 49 (3): 582–600. https://doi.org/10.1017/S0047279419000679.
Chackerian, Richard, and Paul Mavima. 2001. "Comprehensive Administrative Reform Implementation: Moving beyond Single Issue Implementation Research." *Journal of Public Administration Research and Theory* 11 (3): 353–78. https://doi.org/10.1093/oxfordjournals.jpart.a003506.
Choi, Iseul, and Donald Moynihan. 2019. "How to Foster Collaborative Performance Management? Key Factors in the US Federal Agencies." *Public Management Review* 21 (10): 1538–59. https://doi.org/10.1080/14719037.2019.1571275.
Christensen, Julian, Casper M. Dahlmann, Asbjørn H. Mathiasen, Donald P. Moynihan, and Niels B. G. Petersen. 2018. "How Do Elected Officials Evaluate Performance? Goal Preferences, Governance Preferences, and the Process of Goal Reprioritization." *Journal of Public Administration Research and Theory* 28 (2): 197–211. https://doi.org/10.1093/jopart/muy001.
Chun, Young H., and Hal G. Rainey. 2005a. "Goal Ambiguity in U.S. Federal Agencies." *Journal of Public Administration Research and Theory* 15 (1): 1–30. https://doi.org/10.1093/jopart/mui001.
———. 2005b. "Goal Ambiguity and Organizational Performance in U.S. Federal Agencies." *Journal of Public Administration Research and Theory* 15 (4): 529–57. https://doi.org/10.1093/jopart/mui030.
Cohen, Michael D. 2007. "Reading Dewey: Reflections on the Study of Routine." *Organization Studies* 28 (5): 773–86. https://doi.org/10.1177/0170840606077620.
Cook, Brian J. 2014. *Bureaucracy & Self-Government: Reconsidering the Role of Public Administration in American Politics*. 2nd ed. Baltimore, MD: Johns Hopkins University Press.
Corley, Kevin G., Celia V. Harquail, Michael G. Pratt, Mary Ann Glynn, C. Marlene Fiol, and Mary Jo Hatch. 2006. "Guiding Organizational Identity Through Aged Adolescence." *Journal of Management Inquiry* 15 (2): 85–99. https://doi.org/10.1177/1056492605285930.
Council for Excellence in Government, and National Academy for Public Administration. 1998. "Government Performance and Results Act Part 1." *Federal Agencies and Government Waste*. Washington, DC: C-SPAN. https://www.c-span.org/video/?104397-1/government-performance-results-act-part-1.
Croly, Herbert. 1909. *The Promise of American Life*. New York, NY: The Macmillan Company. http://www.gutenberg.org/ebooks/14422.
Cuganesan, Suresh, James Guthrie, and Vedran Vranic. 2014. "The Riskiness of Public Sector Performance Measurement: A Review and Research Agenda." *Financial Accountability & Management* 30 (3): 279–302. https://doi.org/10.1111/faam.12037.
Curry, Dion, Gerhard Hammerschmid, Sebastian Jilke, and Steven Van de Walle. 2015. "Executive Perceptions of Public Sector Reform in Europe." In *International Handbook*

of Public Administration and Governance, edited by A. Massey and K. Johnston, 369–98. Cheltenham, UK: Edward Elgar Publishing. https://www.semanticscholar.org/paper/The-State-and-Perceptions-of-Public-Sector-Reform-Curry-Hammerschmid/508380d33a0b4ec56fdeb374742c008c458b63c6.

De Jong, Maarten. 2016. "Why Agencies Budget for Results – Exploring Institutional Explanations for Performance Budgeting: The Case of Forestry and Air Traffic Control." Dissertation. Rotterdam, Netherlands. hdl.handle.net/1765/79334.

De Lancer Julnes, Patria, and Marc Holzer. 2001. "Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation." *Public Administration Review* 61 (6): 693–708. https://doi.org/10.1111/0033-3352.00140.

De Meur, Gisele, Benoit Rihoux, and Sakura Yamasaki. 2009. "Addressing the Critiques of QCA." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, 147–65. Applied Social Research Methods Series 51. Thousand Oaks, CA: Sage.

De Villiers, Rouxelle. 2017. "QCA in Empirical Marketing Research: An Experiment Featuring Dorah Explorah, Investigating Celebrity Endorsement's Effect on Product Selection." *Australasian Marketing Journal (AMJ)* 25 (3): 225–50. https://doi.org/10.1016/j.ausmj.2017.09.001.

Deephouse, David L., Jonathan Bundy, Leigh P. Tost, and Mark C. Suchman. 2017. "Organizational Legitimacy: Six Key Questions." In *The SAGE Handbook of Organizational Institutionalism*, edited by Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 2:27–54. Thousand Oaks, CA: Sage Publications, Inc.

DeMaio, Carl. 2003. "PART Is Here to Stay Managers Must Take a Hard Look at Program Performance." *Federal Times*, April 14, 2003, sec. NewsBank: Access World News. https://infoweb-newsbankcom. ezproxy.lib.vt.edu/apps/news/document-view?p=AWNB&docref=news/101AD19AD7375A64.

Denhart, Janet V., and Robert B. Denhart. 2015. *The New Public Service: Serving, Not Steering*. 4th ed. New York, NY: Routledge.

Deschamps, Carl, and Jan Mattijs. 2018. "How Organizational Learning Is Supported by Performance Management Systems: Evidence from a Longitudinal Case Study." *Public Performance & Management Review* 41 (3): 469–96. https://doi.org/10.1080/15309576.2018.1462213.

Destler, Katharine N. 2016. "Creating a Performance Culture: Incentives, Climate, and Organizational Change." *American Review of Public Administration* 46 (2): 201–25. https://doi.org/10.1177/0275074014545381.

———. 2017. "A Matter of Trust: Street Level Bureaucrats, Organizational Climate and Performance Management Reform." *Journal of Public Administration Research and Theory* 27 (3): 517–34. https://doi.org/10.1093/jopart/muw055.

Dillman, Don A., Jolene D. Smyth, and Leah M. Christian. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147. https://doi.org/10.2307/2095101.

———. 1991. “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.” In *The New Institutionalism in Organizational Analysis (Electronic Version]*, edited by Walter W. Powell and Paul J. DiMaggio. Chicago, IL: University of Chicago Press.

Doering, Heike, James Downe, Hadar Elraz, and Steve Martin. 2021. “Organizational Identity Threats and Aspirations in Reputation Management.” *Public Management Review* 23 (3): 376–96. https://doi.org/10.1080/14719037.2019.1679234.

Doering, Matthias. 2012. “Der Einfluss von Verwaltungskultur auf die Verwendung von Performance-Daten: Eine Quantitative Untersuchung der Deutschen Kreisfreien Staedte.” Thesis. Potsdam, Germany. http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-57698.

Du Gay, Paul. 1996. “Organizing Identity: Entrepreneurial Governance and Public Engagement.” In *Questions of Cultural Identity*, edited by Stuart Hall and Paul Du Gay. Sage Publications, Inc. https://ebookcentral.proquest.com/lib/vt/detail.action?docID=1474805.

Dubnick, Melvin J., and H. George Frederickson. 2009. “Accountable Agents: Federal Performance Measurement and Third-Party Government.” *Journal of Public Administration Research and Theory* 20 (Supplement 1): i143–59. https://doi.org/10.1093/jopart/mup039.

Dull, Matthew. 2006. “The Politics of Results: Comprehensive Reform and Institutional Choice.” Dissertation, Madison, Wisconsin: University of Wisconsin at Madison. ProQuest.

———. 2009. “Results-Model Reform Leadership: Questions of Credible Commitment.” *Journal of Public Administration Research and Theory: J-PART* 19 (2): 255–84. http://www.jstor.org/stable/29738945.

Durant, Robert F. 2008. “Sharpening a Knife Cleverly: Organizational Change, Policy Paradox, and the ‘Weaponizing’ of Administrative Reforms.” *Public Administration Review* 68 (2): 282–94. https://doi.org/10.1111/j.1540-6210.2007.00862.x.

Ellig, Jerry. 2000. “Performance Report Scorecard: Which Federal Agencies Inform the Public?” Performance Report Scorecard. Fairfax, VA: George Mason University.

Epp, Derek A., and Frank R. Baumgartner. 2017. “Complexity, Capacity, and Budget Punctuations.” *Policy Studies Journal* 45 (2): 247–64. https://doi.org/10.1111/psj.12148.

Executive Office of the President of the United States. 1998. “Government-Wide Performance Plan Fiscal Year 1999.” Washington, DC: Executive Office of the President.

Feldman, Martha S. 2005. “Management and Public Management.” *The Academy of Management Journal* 48 (6): 958–60. https://www.jstor.org/stable/20159716.

Fernandez, Sergio, and Hal G. Rainey. 2006. “Managing Successful Organizational Change in the Public Sector.” *Public Administration Review* 66 (2): 168–76. https://doi.org/10.1111/j.1540-6210.2006.00570.x.

Fischer, Jörn, André Kaiser, and Ingo Rohlfing. 2006. “The Push and Pull of Ministerial Resignations in Germany, 1969–2005.” *West European Politics* 29 (4): 709–35. https://doi.org/10.1080/01402380600842296.

Fischer, Manuel, and Martino Maggetti. 2017. “Qualitative Comparative Analysis and the Study of Policy Processes.” *Journal of Comparative Policy Analysis* 19 (4): 345–61. https://doi.org/10.1080/13876988.2016.1149281.

Fiss, Peer C. 2011. “Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research.” *Academy of Management Journal* 54 (2): 393–420. https://doi.org/10.5465/amj.2011.60263120.
Fiss, Peer C., Mark T. Kennedy, and Gerald F. Davis. 2012. “How Golden Parachutes Unfolded: Diffusion and Variation of a Controversial Practice.” *Organization Science* 23 (4): 1077–99. https://doi.org/10.1287/orsc.1110.0685.
Frederickson, David G., and H. George Frederickson. 2006. *Measuring the Performance of the Hollow State*. Washington, DC: Georgetown University Press.
Frederickson, H. George, Kevin B. Smith, Christopher W. Larimer, and Michael J. Licari. 2016. *The Public Administration Theory Primer*. 3rd ed. Boulder, CO: Westview Press.
Freitag, Markus, and Raphaela Schlicht. 2009. “Educational Federalism in Germany: Foundations of Social Inequality in Education.” *Governance* 22 (1): 47–72. https://doi.org/10.1111/j.1468-0491.2008.01421.x.
Frey, Bruno S., Fabian Homburg, and Margit Osterloh. 2013. “Organizational Control Systems and Pay-for-Performance in the Public Service.” *Organization Studies* 34 (7): 949–72.
Fuenfschilling, Lea, and Bernhard Truffer. 2014. “The Structuration of Socio-Technical Regimes—Conceptual Foundations from Institutional Theory.” *Research Policy* 43 (4): 772–91. https://doi.org/10.1016/j.respol.2013.10.010.
Fuenzalida, Javier, Gregg G. Van Ryzin, and Asmus L. Olsen. 2020. “Are Managers Susceptible to Framing Effects? An Experimental Study of Professional Judgment of Performance Metrics.” *International Public Management Journal*, 1–16. https://doi.org/10.1080/10967494.2020.1752338.
Fukuyama, Francis. 2013. “Commentary: What Is Governance?” *Governance* 26 (3): 347–68. https://doi.org/10.1111/gove.12035.
Fullerton, Andrew S., and Jun Xu. 2012. “The Proportional Odds with Partial Proportionality Constraints Model for Ordinal Response Variables.” *Social Science Research* 41 (1): 182–98. https://doi.org/10.1016/j.ssresearch.2011.09.003.
Furnari, Santi, Donal Crilly, Vilmos F. Misangyi, Thomas Greckhamer, Peer C. Fiss, and Ruth V. Aguilera. 2020. “Capturing Causal Complexity: Heuristics for Configurational Theorizing.” *Academy of Management Review* 46 (4). https://doi.org/10.5465/amr.2019.0298.
Gang Wang, In-Sue Oh, Stephen H. Courtright, and Amy E. Colbert. 2011. “Transformational Leadership and Performance Across Criteria and Levels: A Meta-Analytic Review of 25 Years of Research.” *Group & Organization Management* 36 (2): 223–70. https://doi.org/10.1177/1059601111401017.
Genest-Grégoire, Antoine, Étienne Charbonneau, and Daniel E. Bromberg. 2018. “The Sustainability Assumption in Performance Management Reforms: Revisiting the Patterns of Implementation.” *Public Organization Review* 18 (4): 525–42. https://doi.org/10.1007/s11115-017-0391-5.
George, Bert, Sebastian Desmidt, Poul A. Nielsen, and Martin Baekgaard. 2017. “Rational Planning and Politicians’ Preferences for Spending and Reform: Replication and Extension of a Survey Experiment.” *Public Management Review* 19 (9): 1251–71. https://doi.org/10.1080/14719037.2016.1210905.
George, Bert, and Sanjay K. Pandey. 2017. “We Know the Yin—But Where Is the Yang? Toward a Balanced Approach on Common Source Bias in Public Administration

Scholarship." *Review of Public Personnel Administration* 37 (2): 245–70. https://doi.org/10.1177/0734371X17698189.
Georgellis, Yannis, and Vurain Tabvuma. 2010. "Does Public Service Motivation Adapt?" *Kyklos* 63 (2): 176–91. https://doi.org/10.1111/j.1467-6435.2010.00468.x.
Gerrish, Ed. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76 (1): 48–66. https://doi.org/10.1111/puar.12433.
Gerrits, Lasse M, and Stefan Verweij. 2016. "Qualitative Comparative Analysis as a Method for Evaluating Complex Cases." *Zeitschrift Für Evaluation* 15 (1): 7–22.
Gilad, Sharon. 2015. "Political Pressures, Organizational Identity, and Attention to Tasks: Illustrations from Pre-Crisis Financial Regulation: Regulatory Allocation of Attention." *Public Administration* 93 (3): 593–608. https://doi.org/10.1111/padm.12155.
Gilmour, John B. 2007. "Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrating Budget and Performance." *OECD Journal on Budgeting* 7 (1): 1–40.
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. http://www.jstor.org/stable/3843903.
Gioia, Dennis A., Shubha D. Patvardhan, Aimee L. Hamilton, and Kevin G. Corley. 2013. "Organizational Identity Formation and Change." *Academy of Management Annals* 7 (1): 123–93. https://doi.org/10.5465/19416520.2013.762225.
Glynn, Mary Ann. 2017. "Theorizing the Identity-Institution Relationship: Considering Identity as Antecedent to, Consequence of, and Mechanism for, Processes for Institutional Change." In *The SAGE Handbook of Organizational Institutionalism*, edited by Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 2nd ed., 243–58. Thousand Oaks, CA: Sage Publications, Inc.
Gormley, William T., and Steven J. Balla. 2008. *Bureaucracy and Democracy: Accountability and Performance*. 2nd ed. Washington, DC: CQ Press.
Graeber, David. 2015. *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy*. Brooklyn, NY: Melville House Publishing.
Graham, Erin R., Charles R. Shipan, and Craig Volden. 2013. "The Diffusion of Policy Diffusion Research in Political Science." *British Journal of Political Science* 43 (3): 673–701. https://doi.org/10.1017/S0007123412000415.
Grant Thornton, LLP. 2013. "Taking Measure: Moving from Process to Practice in Performance Management." Washington, DC. https://ourpublicservice.org/publications/taking-measure-moving-from-process-to-practice-in-performance-management/.
Greckhamer, Thomas, Santi Furnari, Peer C. Fiss, and Ruth V. Aguilera. 2018. "Studying Configurations with Qualitative Comparative Analysis: Best Practices in Strategy and Organization Research." *Strategic Organization* 16 (4): 482–95. https://doi.org/10.1177/1476127018786487.
Green, Richard T., and Lawrence Hubbell. 1996. "On Governance and Reinventing Government." In *Refounding Democratic Public Administration: Modern Paradoxes, Postmodern Challenges*, edited by Gary L. Wamsley and James F. Wolf. Thousand Oaks, CA: Sage Publications, Inc.

Greenwood, Royston, C. R. Hinings, and Dave Whetten. 2014. "Rethinking Institutions and Organizations." *Journal of Management Studies*, 1206–20. https://doi.org/10.1111/joms.12070.

Grimmelikhuijsen, Stephan, Sebastian Jilke, Asmus L. Olsen, and Lars Tummers. 2017. "Behavioral Public Administration: Combining Insights from Public Administration and Psychology." *Public Administration Review* 77 (1): 45–56. https://doi.org/10.1111/puar.12609.

Gueorguieva, Vassia, Jean Accius, Carmen Apaza, Lamar Bennett, Clinton Brownley, Shea Cronin, and Panote Preechyanud. 2009. "The Program Assessment Rating Tool and the Government Performance and Results Act: Evaluating Conflicts and Disconnections." *The American Review of Public Administration* 39 (3): 225–45. https://doi.org/10.1177/0275074008319218.

Hall, Peter A., and Rosemary C. R. Taylor. 1996. "Political Science and the Three New Institutionalisms." *Political Studies* 44 (5): 936–57. https://doi.org/10.1111/j.1467-9248.1996.tb00343.x.

Hamilton, Lawrence C. 1992. *Resistant Normality Check and Outlier Identification.* (version 1). Stata Technical Bulletin. University of New Hampshire. https://www.researchgate.net/publication/24136889_Resistant_Normality_Check_and_Outlier_Identification.

Hartmann, Jim, and Anne M. Khademian. 2010. "Culture Change Refined and Revitalized: The Road Show and Guides for Pragmatic Action." *Public Administration Review* 70 (6): 845–56. https://doi.org/10.1111/j.1540-6210.2010.02216.x.

Hatry, Harry P., Elaine Morley, Shelli B. Rossman, and Joseph S. Wholey. 2003. "How Federal Programs Use Outcome Information: Opportunities for Federal Managers." Managing for Results Series. Washington, DC: IBM Endowment for the Business of Government. http://www.businessofgovernment.org/sites/default/files/HatryMorleyRossmanReport.pdf.

Heckman, Jory. 2021. "Biden Undoes Trump Memo Skeptical of Agency Performance Metrics." *Federal News Network*, March 25, 2021. https://federalnewsnetwork.com/agency-oversight/2021/03/biden-undoes-trump-memo-skeptical-of-agency-performance-metrics/.

Heilemann, John. 1995. "The Making of the President 2000." *Wired*, December 1, 1995. https://www.wired.com/1995/12/gorenewt/.

Heinrich, Carolyn J., and Gerald Marschke. 2010. "Incentives and Their Dynamics in Public Sector Performance Management Systems." *Journal of Policy Analysis and Management* 29 (1): 183–208. https://doi.org/10.1002/pam.20484.

Hendren, Kathryn, Qian E. Luo, and Sanjay K. Pandey. 2018. "The State of Mixed Methods Research in Public Administration and Public Policy." *Public Administration Review* 78 (6): 904–16. https://doi.org/10.1111/puar.12981.

Hicklin, Alissa. 2010. "Methods." In *Public Management and Performance: Research Directions*, edited by Richard M. Walker, George A. Boyne, and Gene A. Brewer, 253–66. New York, NY: Cambridge University Press.

Hill, Michael, and Peter Hupe. 2006. "Analysing Policy Processes as Multiple Governance: Accountability in Social Policy." *Policy & Politics* 34 (3): 557–73. https://doi.org/10.1332/030557306777695280.

Ho, Alfred. 2007. "The Governance Challenges of the Government Performance and Results Act: A Case Study of the Substance Abuse and Mental Health Administration." *Public*

Performance & Management Review 30 (3): 369–97. https://doi.org/10.2753/PMR1530-9576300304.

Holm, Jakob M. 2018. "Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance." *Journal of Public Administration Research and Theory* 28 (3): 303–20. https://doi.org/10.1093/jopart/muy017.

Hong, Sounman. 2020. "Performance Management Meets Red Tape: Bounded Rationality, Negativity Bias, and Resource Dependence." *Public Administration Review* 80 (6): 932–45. https://doi.org/10.1111/puar.13213.

Hong, Sounman, Sun Hyoung Kim, and Jieun Son. 2020. "Bounded Rationality, Blame Avoidance, and Political Accountability: How Performance Information Influences Management Quality." *Public Management Review* 22 (8): 1240–63. https://doi.org/10.1080/14719037.2019.1630138.

Hood, Christopher. 2007. "What Happens When Transparency Meets Blame-Avoidance?" *Public Management Review* 9 (2): 191–210. https://doi.org/10.1080/14719030701340275.

———. 2011. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton, NJ: Princeton University Press.

Hopkins, Vincent. 2016. "Institutions, Incentives, and Policy Entrepreneurship." *Policy Studies Journal* 44 (3): 332–48. https://doi.org/10.1111/psj.12132.

Hupe, Peter L., and Michael J. Hill. 2016. "'And the Rest Is Implementation.' Comparing Approaches to What Happens in Policy Processes Beyond Great Expectations." *Public Policy and Administration* 31 (2): 103–21. https://doi.org/10.1177/0952076715598828.

Hvidman, Ulrik, and Simon C. Andersen. 2014. "Impact of Performance Management in Public and Private Organizations." *Journal of Public Administration Research and Theory* 24 (1): 35–58. https://doi.org/10.1093/jopart/mut019.

Jackman, Robert, Juliet Musso, and Christopher Weare. Undated. "The Many Roads to Reform: Identifying Recipes for Success in Performance Management." Unpublished manuscript. https://www.researchgate.net/publication/338558750_The_Many_Roads_to_Reform_ver_20.

Jenkins-Smith, Hank C., Daniel Nohrstedt, Christopher M. Weible, and Paul A. Sabatier. 2014. "The Advocacy Coalition Framework: Foundations, Evolution, and Ongoing Research." In *Theories of the Policy Process*, edited by Paul A. Sabatier and Christopher M. Weible, 3rd ed., 183–223. Boulder, CO: Westview Press.

Jennings, Edward T., and Meg P. Haist. 2004. "Putting Performance Measurement in Context." In *The Art of Governance: Analyzing Management and Administration*, edited by Patricia W. Ingraham and Laurence E. Lynn, 173–94. Washington, DC: Georgetown University Press.

Jepperson, R. 2002. "The Development and Application of Sociological Neoinstitutionalism." In *New Directions in Contemporary Sociological Theory*, edited by J. Berger and M. Zelditch, 229–66. Lanham, MD: Rowman and Littlefield.

Jones, Bryan D., and Frank R. Baumgartner. 2005. *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.

Jung, Chan Su. 2011. "Organizational Goal Ambiguity and Performance: Conceptualization, Measurement, and Relationships." *International Public Management Journal* 14 (2): 193–217. https://doi.org/10.1080/10967494.2011.589760.

———. 2014. “Extending the Theory of Goal Ambiguity to Programs: Examining the Relationship between Goal Ambiguity and Performance.” *Public Administration Review* 74 (2): 205–19. https://doi.org/10.1111/puar.12176.
Kamensky, John. 2018. “Looking Back over My 45 Years of Involvement with Government Reform Efforts.” *Public Administration Review* 78 (2): 305–10. https://doi.org/10.1111/puar.12921.
———. 2021. “Three Parallel Movements to Improve Government Decision Making: Performance, Evidence, and Behavioral Public Administration.” In *The Public Productivity and Performance Handbook*, edited by Marc Holzer and Andrew Ballard, 3rd ed. Philadelphia, PA: Routledge Publishers.
Kamensky, John, and Dan Chenok. 2021. “Lessons from 30 Years of Government Reform Efforts.” *Government Executive*, 2021. https://www.govexec.com/management/2021/03/lessons-30-years-government-reform-efforts/172926/.
Kamensky, John M. 2021. “Performance Management: An Emphasis on Learning.” In *Government Reform: Lessons from the Past for Actions in the Future*, edited by Daniel J. Chenok and John M. Kamensky, 38–42. Washington, DC: IBM Center for the Business of Government. http://www.businessofgovernment.org/report/government-reform-lessons-past-actions-future.
Kettl, Donald. 2009. *The next Government of the United States*. New York, NY: W. W. Norton & Co.
———. 2016. *Escaping Jurassic Government: How to Recover America’s Lost Commitment to Competence*. Washington, DC: Brookings Institution Press.
Khademian, Anne. 1995. “Reinventing a Government Corporation: Professional Priorities and a Clear Bottom Line.” *Public Administration Review* 55 (1): 17. https://doi.org/10.2307/976824.
Kingdon, John. 2003. *Agendas, Alternatives, and Public Policies*. 2nd ed. Longman Classics in Political Science. New York, NY: Addison-Wesley Educational Publishers.
Kodeih, Farah, and Royston Greenwood. 2014. “Responding to Institutional Complexity: The Role of Identity.” *Organization Studies* 35 (1): 7–39. https://doi.org/10.1177/0170840613495333.
Kolenikov, Stas. 2016. “Polychoric, by Any Other ‘Namelist.’” In *2016 Stata Conference 15*. https://www.stata.com/meeting/chicago16/slides/chicago16_kolenikov.pdf.
———. n.d. *Polychoric -- The Polychoric Correlation Package* (version 1.4). Stata 8. University of North Carolina.
Kreiner, Glen. 2011. “Organizational Identity: Culture’s Conceptual Cousin.” In *The Handbook of Organizational Culture and Climate*, edited by Neal M. Ashkenazy, Celeste P. M. Wilderom, and Mark F. Peterson, 2nd ed., 463–80. Thousand Oaks, CA: Sage Publications, Inc.
Kroll, Alexander. 2014. “Why Performance Information Use Varies Among Public Managers: Testing Manager-Related Explanations.” *International Public Management Journal* 17 (2): 174–201. https://doi.org/10.1080/10967494.2014.905409.
———. 2015a. “Explaining the Use of Performance Information by Public Managers: A Planned-Behavior Approach.” *The American Review of Public Administration* 45 (2): 201–15. https://doi.org/10.1177/0275074013486180.

———. 2015b. “Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research.” *Public Performance & Management Review* 38 (3): 459–86. https://doi.org/10.1080/15309576.2015.1006469.
Kroll, Alexander, and Donald P. Moynihan. 2015. “Does Training Matter? Evidence from Performance Management Reforms.” *Public Administration Review* 75 (3): 411–20. https://doi.org/10.1111/puar.12331.
Kroll, Alexander, and Dominik Vogel. 2014. “The PSM-Leadership Fit: A Model of Performance Information Use.” *Public Administration* 92 (4): 974–91. https://doi.org/10.1111/padm.12014.
Lavertu, Stéphane, David E. Lewis, and Donald P. Moynihan. 2013. “Government Reform, Political Ideology, and Administrative Burden: The Case of Performance Management in the Bush Administration.” *Public Administration Review* 73 (6): 845–57. https://doi.org/10.1111/puar.12132.
Lavertu, Stephane, and Donald P. Moynihan. 2013. “Agency Political Ideology and Reform Implementation: Performance Management in the Bush Administration.” *Journal of Public Administration Research and Theory* 23 (3): 521–49. https://doi.org/10.1093/jopart/mus026.
Lee, Chongmyoung, and Richard M. Clerkin. 2017. “Exploring the Use of Outcome Measures in Human Service Nonprofits: Combining Agency, Institutional, and Organizational Capacity Perspectives.” *Public Performance & Management Review* 40 (3): 601–24. https://doi.org/10.1080/15309576.2017.1295872.
Lee, Jung Wook, and Seok Eun Kim. 2012. “Searching for a Strategic Fit: An Empirical Analysis of the Conditions for Performance Management Implementation in U. S. Federal Agencies.” *Public Performance & Management Review* 36 (1): 31–53. https://doi.org/10.2753/PMR1530-9576360102.
Lee, Jung Wook, Hal G. Rainey, and Young Han Chun. 2009. “Of Politics and Purpose: Political Salience and Goal Ambiguity of US Federal Agencies.” *Public Administration* 87 (3): 457–84. https://doi.org/10.1111/j.1467-9299.2009.01772.x.
Liang, Jiaqi. 2014. “Who Maximizes (or Satisfices) in Performance Management? An Empirical Study of the Effects of Motivation-Related Institutional Contexts on Energy Efficiency Policy in China.” *Public Performance & Management Review* 38 (2): 284–315. https://doi.org/10.1080/15309576.2015.983834.
Light, Paul C. 1997. *The Tides of Reform: Making Government Work, 1945-1995*. New Haven, CT: Yale University Press.
Limba, Rekson S., Benny Hutahayan, Solimun Solimun, and Adji Fernandes. 2019. “Sustaining Innovation and Change in Government Sector Organizations: Examining the Nature and Significance of Politics of Organizational Learning.” *Journal of Strategy and Management* 12 (1): 103–15. https://doi.org/10.1108/JSMA-10-2017-0075.
Locke, Edwin A., and Gary P. Latham. 1990. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.
Lounsbury, Michael. 2008. “Institutional Rationality and Practice Variation: New Directions in the Institutional Analysis of Practice.” *Accounting, Organizations and Society* 33 (4–5): 349–61. https://doi.org/10.1016/j.aos.2007.04.001.
Lynn, Laurence E. 1999. “Requiring Bureaucracies to Perform: What Have We Learned from the U.S. Government Performance and Results Act (GPRA)?” *Politics et Management Public* 17 (2): 1–20. https://doi.org/10.3406/pomap.1999.2227.

Lynn, Monty L. 2005. “Organizational Buffering: Managing Boundaries and Cores.” *Organization Studies* 26 (1): 37–61. https://doi.org/10.1177/0170840605046348.
MacIndoe, Heather, and Emily Barman. 2013. “How Organizational Stakeholders Shape Performance Measurement in Nonprofits: Exploring a Multidimensional Measure.” *Nonprofit and Voluntary Sector Quarterly* 42 (4): 716–38. https://doi.org/10.1177/0899764012444351.
Maggetti, Martino, and Yannis Papadopoulos. 2018. “The Principal–Agent Framework and Independent Regulatory Agencies.” *Political Studies Review* 16 (3): 172–83. https://doi.org/10.1177/1478929916664359.
Malatesta, Deanna, and Craig R. Smith. 2014. “Lessons from Resource Dependence Theory for Contemporary Public and Nonprofit Management.” *Public Administration Review* 74 (1): 14–25. https://doi.org/10.1111/puar.12181.
Martins, Luis L. 2005. “A Model of the Effects of Reputational Rankings on Organizational Change.” *Organization Science* 16 (6): 701–20. https://doi.org/10.1287/orsc.1050.0144.
Marx, Axel, Bart Cambre, and Benoit Rihoux. 2013. “Crisp-Set Qualitative Comparative Analysis in Organizational Studies.” In *Configurational Theory and Methods in Organizational Research*, edited by Peer Fiss, Bart Cambre, and Axel Marx, 38:23–47. Emerald Group Publishing. https://www.researchgate.net/publication/236626482_Chapter_2_Crisp-Set_Qualitative_Comparative_Analysis_in_Organizational_Studies_Axel_Marx_Bart_Cambre_Benoit_Rihoux_pp_23_-_47.
Mauro, Sara G., Lino Cinquini, and Giuseppe Grossi. 2018. “External Pressures and Internal Dynamics in the Institutionalization of Performance-Based Budgeting: An Endless Process?” *Public Performance & Management Review* 41 (2): 224–52. https://doi.org/10.1080/15309576.2018.1431137.
McCluskey, E.J. 1956. “Minimization of Boolen Functions.” *The Bell System Technical Journal* 35 (6): 1417–44. https://doi.org/10.1002/j.1538-7305.1956.tb03835.x.
McInnes, Peter, Nic Beech, Linda de Caestecker, Robert MacIntosh, and Michael Ross. 2006. “Identity Dynamics as Barrier to Organizational Change.” *International Journal of Public Administration* 29 (12). https://doi.org/10.1080/01900690600854779.
Meier, Kenneth J., and Laurence J. O’Toole. 2006. “Political Control versus Bureaucratic Values: Reframing the Debate.” *Public Administration Review* 66 (2): 177–92. https://doi.org/10.1111/j.1540-6210.2006.00571.x.
———. 2008. “Management Theory and Occam’s Razor: How Public Organizations Buffer the Environment.” *Administration & Society* 39 (8): 931–58. https://doi.org/10.1177/0095399707309355.
———. 2013. “Subjective Organizational Performance and Measurement Error: Common Source Bias and Spurious Relationships.” *Journal of Public Administration Research and Theory* 23 (2): 429–56. https://doi.org/10.1093/jopart/mus057.
Melkers, Julia, and Katherine Willoughby. 2005. “Models of Performance-Measurement Use in Local Governments: Understanding Budgeting, Communication, and Lasting Effects.” *Public Administration Review* 65 (2): 180–90. https://doi.org/10.1111/j.1540-6210.2005.00443.x.
Mercer, John. 2001. *Statement of John Mercer*. Washington, DC. https://www.john-mercer.com/sites/default/files/library/jm_testimony.pdf.

Metzenbaum, Shelley H. 2009. “Performance Management Recommendations for the New Administration.” Presidential Transition Series. Washington, DC: IBM Center for the Business of Government. http://www.businessofgovernment.org/report/performance-management-recommendations-new-administration.

———. 2013. “Performance Management: The Real Research Challenge.” *Public Administration Review* 73 (6): 857–58. https://doi.org/10.1111/puar.12150.

———. 2015. “Measuring and Improving Government Performance: Learning from Recent U.S. Experience.” In *Managing Under Austerity Delivering Under Pressure: Performance and Productivity in Public Service*, edited by John Wanna, Hsu-Ann Lee, and Sophie Yates, 43–78. Acton, Australia: ANU Press.

Meyer, John W. 2017. “Reflections on Institutional Theories of Organizations.” In *The SAGE Handbook of Organizational Institutionalism*, edited by Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 2nd ed., 831–52. Thousand Oaks, CA: Sage Publications, Inc.

Meyer, John W., and Brian Rowan. 1977. “Institutionalized Organizations: Formal Structure as Myth and Ceremony.” *American Journal of Sociology* 83 (2): 340–63. https://doi.org/10.1086/226550.

Mihm, Christopher. 1999. “Getting Ready for Performance Reporting.” *Public Manager* 28 (3): 7–7. https://link.gale.com/apps/doc/A59045214/AONE?u=anon~653ac048&sid=googleScholar&xid=7b414f07.

Miles, Matthew, and A. Michael Huberman. 1994. *Qualitative Data Analysis*. 2d ed. Thousand Oaks, CA: Sage Publications, Inc.

Miller, Hugh T., and Tansu Demir. 2007. “Policy Communities.” In *Handbook of Public Policy Analysis: Theory, Politics, and Methods*, edited by Frank Fischer, Gerald J. Miller, and Mara S. Sidney. Public Administration and Public Policy 125. Boca Raton, FL: CRC Press.

Moe, Ronald C. 1994. “The ‘Reinventing Government’ Exercise: Misinterpreting the Problem, Misjudging the Consequences.” *Public Administration Review* 54 (2). https://www.jstor.org/stable/976519.

Molina, Anthony D. 2015. “The Virtues of Administration: Values and the Practice of Public Service.” *Administrative Theory & Praxis* 37 (1): 49–69. https://doi.org/10.1080/10841806.2015.999636.

Moore, Mark. 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.

Mortensen, Peter B. 2013. “(De-)Centralisation and Attribution of Blame and Credit.” *Local Government Studies* 39 (2): 163–81. https://doi.org/10.1080/03003930.2012.742015.

Moynihan, Donald P. 2005. “Why and How Do State Governments Adopt and Implement ‘Managing for Results’ Reforms?” *Journal of Public Administration Research and Theory* 15 (2): 219–43. https://doi.org/10.1093/jopart/mui012.

———. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.

———. 2009. “Through A Glass, Darkly: Understanding the Effects of Performance Regimes.” *Public Performance & Management Review* 32 (4): 592–603. https://doi.org/10.2753/PMR1530-9576320409.

———. 2012. “Extra-Network Organizational Reputation and Blame Avoidance in Networks: The Hurricane Katrina Example.” *Governance* 25 (4): 567–88. https://doi.org/10.1111/j.1468-0491.2012.01593.x.

———. 2013. “The New Federal Performance System: Implementing the GPRA Modernization Act.” Improving Performance Series. Washington, DC: IBM Center for the Business of Government. http://www.businessofgovernment.org/sites/default/files/The%20New%20Federal%20Performance%20System.pdf.

Moynihan, Donald P., Sergio Fernandez, Soonhee Kim, Kelly M. LeRoux, Suzanne J. Piotrowski, Bradley E. Wright, and Kaifeng Yang. 2011. “Performance Regimes Amidst Governance Complexity.” *Journal of Public Administration Research and Theory* 21 (Supplement 1): i141–55. https://doi.org/10.1093/jopart/muq059.

Moynihan, Donald P., and Daniel P. Hawes. 2012. “Responsiveness to Reform Values: The Influence of the Environment on Performance Information Use.” *Public Administration Review* 72 (s1): S95–105. https://doi.org/10.1111/j.1540-6210.2012.02653.x.

Moynihan, Donald P., and Patricia W. Ingraham. 2004. “Integrative Leadership in the Public Sector: A Model of Performance-Information Use.” *Administration & Society* 36 (4): 427–53. https://doi.org/10.1177/0095399704266748.

Moynihan, Donald P., and Alexander Kroll. 2016. “Performance Management Routines That Work? An Early Assessment of the GPRA Modernization Act.” *Public Administration Review* 76 (2): 314–23. https://doi.org/10.1111/puar.12434.

Moynihan, Donald P., and Noel Landuyt. 2009. “How Do Public Organizations Learn? Bridging Cultural and Structural Perspectives.” *Public Administration Review* 69 (6): 1097–1105. https://doi.org/10.1111/j.1540-6210.2009.02067.x.

Moynihan, Donald P., and Stéphane Lavertu. 2012. “Does Involvement in Performance Management Routines Encourage Performance Information Use? Evaluating GPRA and PART.” *Public Administration Review* 72 (4): 592–602. https://doi.org/10.1111/j.1540-6210.2011.02539.x.

Moynihan, Donald P., and Sanjay K. Pandey. 2010. “The Big Question for Performance Management: Why Do Managers Use Performance Information?” *Journal of Public Administration Research and Theory* 20 (4). https://doi.org/10.1093/jopart/muq004.

Moynihan, Donald P., Sanjay K. Pandey, and Bradley E. Wright. 2012. “Setting the Table: How Transformational Leadership Fosters Performance Information Use.” *Journal of Public Administration Research and Theory* 22 (1): 143–64. https://doi.org/10.1093/jopart/mur024.

Nam, Taewoo. 2019. “Determinants of Local Public Employee Attitudes toward Government Innovation: Government 3.0 in Korea.” *International Journal of Public Sector Management* 32 (4): 418–34. https://doi.org/10.1108/IJPSM-06-2018-0134.

National Academy of Public Administration. 1999. “Questions and Answers for Improving Government Performance: Operationalizing Performance Management.” Focus. Washington, DC.

National Performance Management Advisory Commission. 2010. “A Performance Management Framework for State and Local Government: From Measurement and Reporting to Management and Improving.” Chicago, IL. https://www.gfoa.org/materials/pmcommission-framework.

National Performance Review. 1993. *From Red Tape to Results: Creating a Government That Works Better and Costs Less*. Washington, DC: U.S. Government Printing Office. HeinOnline.
Neustadt, Richard E., and Ernest R. May. 1986. *Thinking in Time: The Uses of History for Decision Makers*. New York, NY: The Free Press.
Newcomer, Kathryn, and Clint Brass. 2021. "Approaches to Improving Performance in Government: Making Sense of Where We've Been and What's Next." In *The Public Productivity and Performance Handbook*, edited by Marc Holzer and Andrew Ballard, 3rd ed., 48–67. Washington, DC: Routledge.
Nielsen, Poul A. 2014. "Performance Management, Managerial Authority, and Public Service Performance." *Journal of Public Administration Research and Theory* 24 (2): 431–58. https://doi.org/10.1093/jopart/mut025.
Nielsen, Poul A, and Martin Baekgaard. 2015. "Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment." *Journal of Public Administration Research and Theory* 25 (2): 545–69. https://doi.org/10.1093/jopart/mut051.
Nielsen, Poul A., and Donald P. Moynihan. 2017. "How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy." *Journal of Public Administration Research and Theory* 27 (2): 269–83. https://doi.org/10.1093/jopart/muw060.
Niskanen, William. 1994. *Bureaucracy and Public Economics*. Brookfield, VT: Edward Elgar Publishing.
Nitzl, Christian, Maria F. Sicilia, and Ileana Steccolini. 2019. "Exploring the Links Between Different Performance Information Uses, NPM Cultural Orientation, and Organizational Performance in the Public Sector." *Public Management Review* 21 (5): 686–710. https://doi.org/10.1080/14719037.2018.1508609.
O'Brien, Robert M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality and Quantity* 41: 673–390. https://doi.org/10.1007/s11135-006-9018-6.
Ohemeng, Frank L. K. 2011. "Institutionalizing the Performance Management System in Public Organizations in Ghana: Chasing a Mirage?" *Public Performance & Management Review* 34 (4): 467–88. https://doi.org/10.2753/PMR1530-9576340402.
Olsen, Asmus L. 2017. "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." *Journal of Public Administration Research and Theory* 27 (4): 562–80. https://doi.org/10.1093/jopart/mux023.
Organisation for Economic Co-Operation and Development. 1997. "In Search of Results: Performance Management Practices." OECD. https://www.oecd.org/unitedstates/36144694.pdf.
Osborne, David E., and Ted Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*. Reading, MA: Addison-Wesley.
Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York, NY: Cambridge University Press.
———. 2011. "Background on the Institutional Analysis and Development Framework." *Policy Studies Journal* 39 (1): 7–27. https://doi.org/10.1111/j.1541-0072.2010.00394.x.
O'Toole, Laurence J., and Kenneth J. Meier. 2011. *Public Management: Organizations, Governance, and Performance*. New York, NY: Cambridge University Press.

———. 2015. “Public Management, Context, and Performance: In Quest of a More General Theory.” *Journal of Public Administration Research and Theory* 25 (1): 237–56. https://doi.org/10.1093/jopart/muu011.
Ozturk, Oguzhan. 2021. “Bibliometric Review of Resource Dependence Theory Literature: An Overview.” *Management Review Quarterly* 71: 525–52. https://doi.org/10.1007/s11301-020-00192-8.
Pfeffer, Jeffrey. 2006. “A Review of: ‘Like Ships Passing in the Night: The Separate Literatures of Organization Theory and Public Management.’” *International Public Management Journal* 9 (4): 457–65. https://doi.org/10.1080/10967490601077426.
Pfeffer, Jeffrey, and Gerald R. Salancik. 2003. *The External Control of Organizations: A Resource Dependence Perspective*. Stanford, CA: Stanford University Press.
Pfiffner, Roger. 2019. “Why Performance Information Use Requires a Managerial Identity: Evidence from the Field of Human Services.” *Public Performance & Management Review* 42 (2): 405–31. https://doi.org/10.1080/15309576.2018.1464936.
Pollitt, Christopher. 2006. “Performance Management in Practice: A Comparative Study of Executive Agencies.” *Journal of Public Administration Research and Theory* 16 (1): 25–44. https://doi.org/10.1093/jopart/mui045.
———. 2013. “The Logics of Performance Management.” *Evaluation* 19 (4): 346–63. https://doi.org/10.1177/1356389013505040.
Poole, Ed G. 2019. “How Institutional Culture Trumps Tier Effects: Evidence from Government Responsiveness to FOI Requests.” *Journal of Public Administration Research and Theory* 29 (2): 210–26. https://doi.org/10.1093/jopart/muy039.
Powell, Walter W., and Claus Rerup. 2017. “Opening the Black Box: Microfoundations of Institutions.” In *The SAGE Handbook of Organizational Institutionalism*, edited by Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 2nd ed., 311–37. Thousand Oaks, CA: Sage Publications, Inc.
President of the United States. 1989. “Budget of the United States Government Fiscal Year 1990.” Government Printing Office. https://fraser.stlouisfed.org/files/docs/publications/usbudget/bus_1990.pdf.
Proeller, Isabella, Alexander Kroll, and Anne-Kathrin Meier. 2012. “Performance Management in Den Kreisfreien Staedten: Einschaetzungen Aus Perspektive Der Mittleren Fuehrungskraefte.” *Verwaltung Und Management* 18 (4): 180–87. https://www.uni-potsdam.de/fileadmin/projects/ls-puma/Performance_Management_in_kreisfreien_St%C3%A4dten.pdf.
Radin, Beryl A. 2000. “The Government Performance and Results Act and the Tradition of Federal Management Reform: Square Pegs in Round Holes?” *Journal of Public Administration Research and Theory* 10 (1): 111–35. https://doi.org/10.1093/oxfordjournals.jpart.a024260.
———. 2006. *Challenging the Performance Movement: Accountability, Complexity and Democratic Values*. Washington, DC: Georgetown University Press.
Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of California Press.
———. 2009. “Qualitative Comparative Analysis Using Fuzzy Sets (FSQCA).” In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Relation Techniques*, edited by Benoit Rihoux and Charles C. Ragin, 51:87–121. Applied Social Research Methods Series. Thousand Oaks, CA: Sage Publications, Inc.

———. 2014. *The Comparative Method*. 2nd ed. Oakland, CA: University of California Press.
Raines, Franklin D. 1997. "The Government Performance and Results Act: Report to the President and Congress from the Director of the Office of Management and Budget." Washington, DC: Office of Management and Budget. https://govinfo.library.unt.edu/npr/library/news/gprarpt.html#decision_designate.
Rainey, Hal G. 2014. *Understanding and Managing Public Organizations*. 5th ed. San Francisco, CA: Jossey-Bass.
Rainey, Hal G., and Chan Su Jung. 2010. "Extending Goal Ambiguity Research in Government: From Organizational Goal Ambiguity to Programme Goal Ambiguity." In *Public Management and Performance: Research Directions*, edited by Richard M. Walker, George A. Boyne, and Gene A. Brewer, 34–59. New York, NY: Cambridge University Press.
Reay, Trish, Samia Chreim, Karen Golden-Biddle, Elizabeth Goodrick, B. E. Bernie Williams, Ann Casebeer, Amy Pablo, and C. R. Bob Hinings. 2013. "Transforming New Ideas into Practice: An Activity Based Perspective on the Institutionalization of Practices: Transforming New Ideas into Practice." *Journal of Management Studies* 50 (6): 963–90. https://doi.org/10.1111/joms.12039.
Rihoux, Benoit, and Gisele De Meur. 2009. "Crisp-Set Qualitative Comparative Analysis (CsQCA)." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, 51:33–68. Applied Social Research Methods Series. Thousand Oaks, CA: Sage.
Rihoux, Benoit, Charles C. Ragin, Sakura Yamasaki, and Damian Bol. 2009. "Conclusions-The Way(s) Ahead." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, 51:167–77. Applied Social Research Methods Series. Thousand Oaks, CA: Sage Publications, Inc.
Rohr, John A. 2002. *Civil Servants and Their Constitutions*. Lawrence, KA: University Press of Kansas.
Romzek, Barbara, and Melvin Dubnick. 1987. "Accountability in the Public Sector: Lessons from the Challenger Tragedy." *Public Administration Review* 47: 227–38. https://www.jstor.org/stable/975901.
Ryu, Jay E. 2011. *Bounded Bureaucracy and the Budgetary Process in the United States*. New Brunswick, NJ: Transaction Publishers.
Sabatier, Paul A. 1986. "Top-Down and Bottom-Up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis." *Journal of Public Policy* 6 (1): 21–48. https://doi.org/10.1017/S0143814X00003846.
Sager, Fritz, and Eva Thomann. 2017. "Multiple Streams in Member State Implementation: Politics, Problem Construction and Policy Paths in Swiss Asylum Policy." *Journal of Public Policy* 37 (3): 287–314. https://doi.org/10.1017/S0143814X1600009X.
Sarta, Andrew, Rodolphe Durand, and Jean-Philippe Vergne. 2021. "Organizational Adaptation." *Journal of Management* 47 (1): 43–75. https://doi.org/10.1177/0149206320929088.
Schachter, Hindy L. 2002. "Women, Progressive-Era Reform, and Scientific Management." *Administration & Society* 34 (5): 563–78. https://doi.org/10.1177/009539902237276.
———. 2007. "Does Frederick Taylor's Ghost Still Haunt the Halls of Government? A Look at the Concept of Governmental Efficiency in Our Time." *Public Administration Review* 67 (5): 800–810. https://doi.org/10.1111/j.1540-6210.2007.00768.x.

Schillemans, Thomas, and Madalina Busuioc. 2015. “Predicting Public Sector Accountability: From Agency Drift to Forum Drift.” *Journal of Public Administration Research and Theory* 25 (1): 191–215. https://doi.org/10.1093/jopart/muu024.

Schneider, Carsten Q., and Claudius Wagemann. 2013. *Set-Theoretic Methods for the Social Sciences*. Cambridge, United Kingdom: Cambridge University Press.

Schofer, Evan. 2007. “Multilevel Models 3.” Class Presentation, Irvine, CA. https://slidetodoc.com/multilevel-models-3-sociology-8811-class-25-copyright/.

Schwartz-Soicher, Ofira. 2017. “Multiple Imputation of MIssing Data Using Stata.” Class Presentation, Princeton, NJ. https://dss.princeton.edu/training/MIStata.pdf.

Scott, W. Richard. 2001. *Institutions and Organizations*. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc.

———. 2017. “Institutional Theory: Onward and Upward.” In *The SAGE Handbook of Organizational Institutionalism*, edited by Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 2nd ed., 853–71. Thousand Oaks, CA: Sage Publications, Inc.

Scott, W. Richard, and Gerald F. Davis. 2007. *Organizations and Organizing: Rational, Natural, and Open Systems Perspectives*. Upper Saddle River, NJ: Pearson Prentice Hall.

Selznick, Philip. 1957. *Leadership in Administration: A Sociological Interpretation*. Berkeley, CA: University of California Press.

Shand, David. 2007. “Performance Auditing and Performance Budgeting.” In *Performance Budgeting: Linking Results and Funding*, edited by Marc Robinson, 88–109. New York, NY: Palgrave Macmillan.

Simon, Herbert A. 1997. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. 4th ed. New York, NY: The Free Press.

Skowronek, Stephen. 1984. *Building a New American State: The Expansion of National Administrative Capacities, 1877-1920*. New York, NY: Cambridge University Press.

Smets, Michael, Angela Aristidou, and Richard Whittington. 2017. “Towards a Practice-Driven Institutionalism.” In *The SAGE Handbook of Organizational Institutionalism*, edited by Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 365–91. Thousand Oaks, CA: Sage Publications, Inc.

Smets, Michael, Paula Jarzabkowski, Gary T. Burke, and Paul Spee. 2015. “Reinsurance Trading in Lloyd’s of London: Balancing Conflicting-yet-Complementary Logics in Practice.” *Academy of Management Journal* 58 (3): 932–70. https://doi.org/10.5465/amj.2012.0638.

Spekle, Roland F., and Frank H. M. Verbeeten. 2014. “The Use of Performance Measurement Systems in the Public Sector: Effects on Performance.” *Management Accounting Research* 25: 131–46. https://doi.org/10.1016/j.mar.2013.07.004.

Stanley, J. Woody, and Ryan Lutz. 2021. “Implementation of the Federal Performance Framework under Presidents Obama and Trump: A Comparative Analysis of Agency Strategic Plans.” *Public Performance & Management Review*, 1–24. https://doi.org/10.1080/15309576.2021.1884577.

Stivers, Camilla. 1995. “Settlement Women and Bureau Men: Constructing a Usable Past for Public Administration.” *Public Administration Review* 55 (6): 522–29. https://www.jstor.org/stable/3110343.

Tabachnik, Barbara G, and Linda S. Fidell. 2007. *Using Multivariate Statistics*. 5th ed. Allyn and Beacon/Pearson Education.

Taylor, Jeannette. 2011. “Factors Influencing the Use of Performance Information for Decision Making in Australian State Agencies.” *Public Administration* 89 (4): 1316–34. https://doi.org/10.1111/j.1467-9299.2011.02008.x.

Teodoro, Manuel P. 2014. “When Professionals Lead: Executive Management, Normative Isomorphism, and Policy Implementation.” *Journal of Public Administration Research and Theory* 24 (4): 983–1004. https://doi.org/10.1093/jopart/muu039.

Terry, Larry. 2003. *Leadership of Public Bureaucracies*. 2nd ed. New York, NY: Routledge.

Thaler, Richard H. 2015. *Misbehaving: The Making of Behavioral Economics*. New York, NY: W.W. Norton and Co.

Thiem, Alrik. 2017. “Conducting Configurational Comparative Research With Qualitative Comparative Analysis: A Hands-On Tutorial for Applied Evaluation Scholars and Practitioners.” *American Journal of Evaluation* 38 (3): 420–33. https://doi.org/10.1177/1098214016673902.

Thiem, Alrik, L. Mkrtchyan, T. Hasebrouck, and D. Sanchez. 2020. “Algorithmic Bias in Social Research: A Meta-Analysis.” Edited by Sandro Meloni. *PLoS ONE* 15 (6): e0233625. https://doi.org/10.1371/journal. pone.0233625.

Thomann, Eva. 2015. “Is Output Performance All About the Resources? A Fuzzy-Set Qualitative Comparative Analysis of Street-Level Bureaucrats in Switzerland.” *Public Administration* 93 (1): 177–94. https://doi.org/10.1111/padm.12130.

Thomann, Eva, and Joern Ege. 2020. “Qualitative Comparative Analysis (QCA) in Public Administration.” In *Oxford Research Encyclopedias*. https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-1444.

Thomann, Eva, and Martino Maggetti. 2020. “Designing Research with Qualitative Comparative Analysis (QCA): Approaches, Challenges, and Tools.” *Sociological Methods & Research* 49 (2): 356–86. https://doi.org/10.1177/0049124117729700.

Thompson, Fred. 2001. “Government at the Brink: Urgent Federal Government Management Problems Facing the Bush Administration (Volume 1).” Washington, DC: Committee on Governmental Affais, United States Senate.

Thompson, James D. 1967. *Organizations in Action: Social Science Bases of Administrative Theory*. New York, NY: McGraw-Hill.

Tolbert, Pamela A., and Lynne G. Zucker. 1996. “The Institutionalization of Institutional Theory.” In *Handbook of Organization Studies [Electronic Version]*, edited by Stewart R. Clegg, Cynthia Hardy, and Walter R. Nord, 175–90. Thousand Oaks, CA: Sage Publications, Inc.

Tolbert, Pamela S., and Lynne G. Zucker. 1983. “Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform, 1880-1935.” *Administrative Science Quarterly* 28 (1): 22. https://doi.org/10.2307/2392383.

Tummers, Lars, Bram Steijn, and Victor Bekkers. 2012. “Explaining the Willingness of Public Professionals to Implement Public Policies: Content, Context, and Personality Characteristics.” *Public Administration* 90 (3): 716–36. https://doi.org/10.1111/j.1467-9299.2011.02016.x.

Turner, Scott F., and Violina Rindova. 2012. “A Balancing Act: How Organizations Pursue Consistency in Routine Functioning in the Face of Ongoing Change.” *Organization Science* 23 (1): 24–46. https://doi.org/10.1287/orsc.1110.0653.

University of California at Santa Barbara. The American Presidency Project. n.d. "1992 Democratic Party Platform (July 13, 1992)." Accessed April 15, 2021. https://www.presidency.ucsb.edu/documents/1992-democratic-party-platform.

U.S. Congress. House. Committee on Oversight and Government Reform. 2015. *Oversight Plans for All House Committees*. 114th Cong. 1st Sess., 2015, H. Rep. 114-82. https://www.govinfo.gov/content/pkg/CRPT-114hrpt82/pdf/CRPT-114hrpt82.pdf.

U.S. Congress. House. Committee on Oversight and Government Reform. Subcommittee on Government Management, Organization, and Procurement. 2009. *Oversight of Federal Financial Management*. 111th Cong., 1st Sess., 2009, H. Rep. 111-57. https://www.govinfo.gov/content/pkg/CHRG-111hhrg57382/pdf/CHRG-111hhrg57382.pdf.

U.S. Congress. Senate. Committee on Appropriations. 2000. *Special Report of the Committee on Appropriations United States Senate on the Government Performance and Results Act of 1993, July 18, 2000*. 106th Cong., 2d Sess., 2000, S. Rep. 106-347. Washington: U.S. Government Printing Office. https://www.congress.gov/106/crpt/srpt347/CRPT-106srpt347.pdf.

U.S. Congress. Senate. Committee on the Budget. 2009. *Performance-Informed Budgeting: Opportunities to Reduce Cost and Improve Service*. 111th Cong., 1st Sess., 2009, S. Hrg. 111-381. https://www.govinfo.gov/app/details/CHRG-111shrg55507/CHRG-111shrg55507.

U.S. Congress. House. Subcommittee on Government Efficiency, Financial Management, and Intergovernmental Relations. 2001. *The Results Act: Has It Met Congressional Expectations?* 107th Cong., 1st Sess., 2001, H. Hrg. 107-75. https://www.govinfo.gov/content/pkg/CHRG-107hhrg80373/pdf/CHRG-107hhrg80373.pdf.

U.S. Congress. Senate. Committee on Homeland Security and Government Affairs. Federal Financial Management, Government Information, Federal Services, and International Security Subcommittee. 2009. *Getting to Better Government: Focusing on Performance*. 111th Cong., 1st Sess., 2009. S. Hrg. 111-633. https://www.govinfo.gov/content/pkg/CHRG-111shrg53843/pdf/CHRG-111shrg53843.pdf.

U.S. Congress. Senate. Committee on the Budget. 2013. *Silo Busting; Effective Strategies for Government Reorganization*. 113th Cong., 1st Sess., 2013, S. Hrg. 113-231. https://www.govinfo.gov/content/pkg/CHRG-113shrg85904/pdf/CHRG-113shrg85904.pdf.

U.S. Congress. Senate. Committees on Governmental Affairs and Appropriations. 1997. *Implementation of the Government Performance and Results Act of 1993*. 105th Cong., 1st Sess. 1997. S. Hrg. 105-198. HeinOnline.

U.S. Congress. Senate. Committee on Governmental Affairs. 1993. "Government Performance and Results Act of 1993." Conference Report 103–58. Washington, DC: United States Senate. https://obamawhitehouse.archives.gov/omb/mgmt-gpra/gprptm.

U.S. Congress. Senate. Committee on Homeland Security and Governmental Affairs. 2011. *Roadmap for a More Efficient and Accountable Federal Government: Implementing the GPRA Modernization Act*. 112th Cong., 1st Sess., 2011, S. Hrg. 112-156. https://www.congress.gov/112/chrg/CHRG-112shrg67636/CHRG-112shrg67636.pdf.

U.S. Department of Defense. 2020. *FY 2021 Annual Performance Plan and FY 2019 Annual Performance Report*. Washington, DC: U.S. Department of Defense. https://cmo.defense.gov/Publications/Annual-Performance-Plan-and-Performance-Report/.

U.S. General Accounting Office. 1992a. "Government Management Issues." GAO/OCG-93-3TR. Washington, DC: U.S. General Accounting Office.

———. 1992b. "Performance Measurement: An Important Tool in Managing for Results." GAO/TGGD-92-35. Washington, DC: U.S. General Accounting Office.

———. 1992c. "Program Performance Measures: Federal Agency Collection and Use of Performance Data." GAO/GGD-92-65. Washington, DC: U.S. General Accounting Office.

———. 1997a. "GPRA: Managerial Accountability and Flexibility Pilot Did Not Work as Intended." GAO/GGD-97-36. Washington, DC: U.S. General Accounting Office.

———. 1997b. "Managing for Results: Analytic Challenges in Measuring Performance." GAO/HEHS/GGD-97-138. Washington, DC: U.S. General Accounting Office.

———. 1997c. "Managing for Results: Enhancing the Usefulness of GPRA Consultations Between Executive Branch and Congress." GAO/T-GGD-97-56. Washington, DC: U.S. General Accounting Office.

———. 1997d. "The Government Performance and Results Act: 1997 Governmentwide Implementation Will Be Uneven." GGD-97-109. https://www.gao.gov/products/GGD-97-109.

———. 1998. "The Results Act: Assessment of the Governmentwide Performance Plan for Fiscal Year 1999." GAO/AIMD/GGD-98-159. Washington, DC: U.S. General Accounting Office.

———. 2000a. "Determining Performance and Accountability Challenges and High Risks." GAO-01-159SP. Washington, DC. https://www.gao.gov/assets/210/200448.pdf.

———. 2000b. "Managing for Results: Challenges Agencies Face in Producing Credible Performance Information." GAO/GGD-00-52. Washington, DC: U.S. General Accounting Office.

———. 2000c. "Managing for Results: Federal Managers' Views Show Need for Ensuring Top Leadership Skills." GAO-01-127. https://www.gao.gov/assets/gao-01-127.pdf.

———. 2000d. "Managing for Results: Continuing Challenges to Effective GPRA Implementation." GAO/T-GGD-DD-178. https://www.gao.gov/assets/t-ggd-00-178.pdf.

———. 2001. "Managing for Results: Federal Managers' Views on Key Management Issues Vary Widely Across Agencies." GAO-01-592. Washington, DC: U.S. General Accounting Office.

———. 2004a. "Performance Budgeting: Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget." GAO-04-174. https://www.gao.gov/assets/gao-04-174.pdf.

———. 2004b. "Results-Oriented Government: GPRA Has Established a Solid Foundation for Achieving Greater Results." GAO 04-38. https://www.gao.gov/products/gao-04-38.

U.S. Government Accountability Office. 2005. "Program Evaluation: OMB's PART Reviews Increased Agencies' Attention to Improving Evidence of Program Results." Report GAO-06-67. Washington, DC: U.S. Government Accountability Office.

———. 2006. “Human Capital: Aligning Senior Executives’ Performance with Organizational Results Is an Important Step Towards Governmentwide Transformation.” Testimony GAO-06-1125T. https://www.gao.gov/assets/gao-06-1125t.pdf.
———. 2008. “Government Performance: Lessons Learned for the Next Administration on Using Performance Information to Improve Results.” GAO-08-1026T. https://www.gao.gov/assets/gao-08-1026t.pdf.
———. 2011a. “Government Performance: GPRA Modernization Act Provides Opportunities to Help Address Fiscal, Performance, and Management Challenges.” GAO-11-466T. https://www.gao.gov/new.items/d11466t.pdf.
———. 2011b. “Streamlining Government: Key Practices from Select Efficiency Initiatives Should Be Shared Governmentwide.” GAO-11-908. Washington, DC: U.S. Government Accountability Office. https://www.gao.gov/assets/gao-11-908.pdf.
———. 2013a. “Managing for Results: 2013 Federal Managers Survey on Organizational Performance and Management Issues.” GAO-13-519SP. https://files.gao.gov/special.pubs/gao-13-519sp/results.htm#question_72.
———. 2013b. “Managing for Results: Agencies Have Elevated Performance Management Leadership Roles, but Additional Training Is Needed.” GAO-13-356. https://www.gao.gov/assets/gao-13-356.pdf.
———. 2015. “Managing for Results: Implementation of GPRA Modernization Act Has Yielded Mixed Progress in Addressing Pressing Governance Challenges.” GAO-15-819. https://www.gao.gov/assets/gao-15-819.pdf.
———. 2016. “Department of Housing and Urban Development: Actions Needed to Incorporate Key Practices into Management Functions and Program Oversight.” GAO-16-497. Washington, DC: U.S. Government Accountability Office. https://www.gao.gov/assets/gao-16-497.pdf.
———. 2017a. “Managing for Results: Further Progress Made in Implementing the GPRA Modernization Act, but Additional Actions Needed to Address Pressing Governance Challenges.” GAO-17-775. https://www.gao.gov/products/GAO-17-775.
———. 2017b. “Survey of Organizational Performance and Management Issues, 2017.” https://files.gao.gov/special.pubs/gao-17-776sp/resultsall.htm.
———. 2017c. “Survey of Organizational Performance and Management Issues, 2017 [Data Set].”
———. 2018. “Managing for Results: Government-Wide Actions Needed to Improve Agencies’ Use of Performance Information in Decision Making.” GAO-18-609SP. https://www.gao.gov/assets/700/694269.pdf.
U.S. Library of Congress, Congressional Research Service. 2004. *The Bush Administration’s Program Assessment Rating Tool*. By Clinton T. Brass, RL32663 (2014). https://crsreports.congress.gov/product/details?prodcode=RL32663.
———. 2005. *Performance Management and Budgeting in the Federal Government: Brief History and Recent Developments*. By Virginia A. McMurtry. RL32164 (2005). https://crsreports.congress.gov/product/details?prodcode=RL32164.
U.S. Office of Management and Budget. 2003. “Circular No. A-11: Preparation, Submission, and Execution of the Budget.” Circular. Washington, DC. https://georgewbush-whitehouse.archives.gov/omb/circulars/a11/current_year/a_11_2003.pdf.
———. 2011. *Delivering an Efficient, Effective, and Accountable Government*. By Jacob J. Lew, OMB M-11-31. Washington, DC.

https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2011/m11-31.pdf.
U.S. Office of Personnel Management. 2019. "Federal Employment Viewpoint Survey: Governmentwide Management Report." Washington, DC. https://www.opm.gov/fevs/reports/governmentwide-reports/governmentwide-management-report/governmentwide-report/2019/2019-governmentwide-management-report.pdf.
U.S. President. Executive Order. "Improving Government Program Performance, Executive Order 13450 of November 13, 2007." *Federal Register* 72 (220): 64517–22. https://www.govinfo.gov/content/pkg/FR-2007-11-15/pdf/07-5726.pdf.
Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. 2nd ed. Interdisciplinary Statistics Series. Boca Raton, FL: Chapman and Hall/CRC. https://stefvanbuuren.name/fimd/.
Van Dooren, Wouter. 2005. "What Makes Organisations Measure? Hypotheses on the Causes and Conditions for Performance Measurement." *Financial Accountability and Management* 21 (3): 363–83. https://doi.org/10.1111/j.0267-4424.2005.00225.x.
Van Thiel, Sandra, and Frans L. Leeuw. 2002. "The Performance Paradox in the Public Sector." *Public Performance & Management Review* 25 (3): 267–81. https://www.jstor.org/stable/3381236.
Van Wart, Montgomery. 2003. "Public-Sector Leadership Theory: An Assessment." *Public Administration Review* 63 (2): 214–28. http://www.jstor.org/stable/977592.
———. 2012. *Leadership in Public Organizations: An Introduction*. 2nd ed. Armonk, NY: M. E. Sharpe.
———. 2013. "Administrative Leadership Theory: A Reassessment after Ten Years." *Public Administration*, 521–43. https://doi.org/10.1111/padm.12017.
Vogel, Rick. 2014. "What Happened to the Public Organization? A Bibliometric Analysis of Public Administration and Organization Studies." *The American Review of Public Administration* 44 (4): 383–408. https://doi.org/10.1177/0275074012470867.
Von Hippel, Paul T. 2020. "How Many Imputations Do You Need? A Two-Stage Calculation Using a Quadratic Rule." *Sociological Methods & Research* 49 (3): 699–718. https://doi.org/10.1177/0049124117747303.
Waggoner, Daniel B., Andy D. Neely, and Mike P. Kennerley. 1999. "The Forces That Shape Organisational Performance Measurement Systems: An Interdisciplinary Review." *International Journal of Production Economics*, no. 60: 53–60. https://doi.org/10.1016/S0925-5273(98)00201-1.
Walker, Richard M., Chan Su Jung, and George A. Boyne. 2013. "Marching to Different Drummers? The Performance Effects of Alignment between Political and Managerial Perceptions of Performance Management." *Public Administration Review* 73 (6): 833–44. https://doi.org/10.1111/puar.12131.
Weick, Karl E. 1993. "The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster." *Administrative Science Quarterly* 38 (4): 628–52. https://doi.org/10.2307/2393339.
Weinstock, Matthew. 2003. "Under the Microscope." *Government Executive* 35 (1): 37–40.
Wildavsky, Aaron, and Naomi Caiden. 2004. *The New Politics of the Budgetary Process*. 5th ed. New York, NY: Pearson Longman.
Williams, Daniel W. 2002. "Before Performance Measurement." *Administrative Theory & Praxis* 24 (3): 457–86. https://doi.org/10.1080/10841806.2002.11029377.

Williams, Richard. 2006. “Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables.” *The Stata Journal: Promoting Communications on Statistics and Stata* 6: 58–82. https://doi.org/10.1177/1536867X0600600104.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. 2nd ed. United States of America: Basic Books, Inc.

Wilson, Woodrow. 1887. “The Study of Administration.” *Political Science Quarterly* II (2): 197–222. http://www.jstor.org/stable/2139277.

Wood, Abby K., and David E. Lewis. 2017. “Agency Performance Challenges and Agency Politicization.” *Journal of Public Administration Research and Theory* 27 (4): 581–95. https://doi.org/10.1093/jopart/mux014.

Yamasaki, Sakura, and Benoit Rihoux. 2009. “A Commented Review of Applications.” In *Configurational Comparative Methods: Qualitative Comparative Analysis and Related Techniques*, 51:123–45. Applied Social Research Methods Series. Thousand Oaks, CA: Sage Publications, Inc.

Yang, Kaifeng. 2011. “The Sisyphean Fate of Government-Wide Performance Accountability Reforms: Federal Performance Management Efforts and Employees’ Daily Work, 2002-2008.” *Public Performance & Management Review* 35 (1): 149–76. https://doi.org/10.2753/PMR1530-9576350108.

Yang, Kaifeng, and Jun Yi Hsieh. 2007. “Managerial Effectiveness of Government Performance Measurement: Testing a Middle-Range Model.” *Public Administration Review* 67 (5): 861–79. https://doi.org/10.1111/j.1540-6210.2007.00774.x.

Yang, Kaifeng, and Sanjay K. Pandey. 2009. “How Do Perceived Political Environment and Administrative Reform Affect Employee Commitment?” *Journal of Public Administration Research and Theory* 19 (2): 335–60. https://doi.org/10.1093/jopart/mun002.

Yetano, Ana. 2013. “What Drives the Institutionalization of Performance Measurement and Management in Local Government?” *Public Performance & Management Review* 37 (1): 59–86. https://doi.org/10.2753/PMR1530-9576370103.

Yin, Robert K. 2014. *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage.

Young, Shalanda D. 2021. “Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans [M-21-27 Memo to Heads of Executive Departments and Agencies].” Office of Management and Budget. https://www.whitehouse.gov/wp-content/uploads/2021/06/M-21-27.pdf.

Appendices

Appendix A. Comparison of GPRA, PART, and GPRAMA Performance Regimes

Products & Processes	GPRA	PART	GPRAMA
Executive Branch-Wide Products and Processes			
Goal Setting	None	None	Four-year Federal Government Performance Goals (FGPG), aka "Cross-Agency Priority" (CAP) goals, a limited number of policy outcome and management improvement goals (1)
Performance Plans	None	None	Biannual Federal Government Performance Plans (FGPP). Per OMB A-11 part 6, the FGPP defines CAP goal targets, action plan, goal leaders and contributing programs. (2)
Agency Level Products and Processes			
Strategic Plans	Strategic Plans (5-year) with 3-year review cycle	Continued GPRA	4-Year Agency Strategic Plans (ASP) with 4-year review cycle including strategic goals and objectives.
Performance Plans	Performance Plans (annual) covering program activity in agency budget. Consistent with Strategic Plans.	"Performance Budgets" (OMB) (annual, starting in FY 2005)	Agency Performance Plans (APP) (annual) which include agency performance goals for current and next FY. APPs must be consistent with ASPs and describe how they contribute to ASP goals and CAP goals.
Priority Goals	Only goals in performance plan	Only goals in performance plan	Agency Priority Goals (APG), subset of APP goals (bi-annual). Usually 2-8 APGs (per OMB A-11). Informed by CAP goals, with quarterly milestones.
Performance Reports	Program Performance Reports (annual)	Continued GPRA	Agency Performance Updates (APU) or Performance and Accountability Reports (PAR) (3)
Program Evaluations	Not explicitly required	Core agency activity	Not explicitly required, informs ASP goals and objectives
Accountability Mechanisms			
Congressional Oversight	Strategic plans report agency goals, objectives, how achieved, and how performance plan goals relate to strategic plan; program performance reports communicate agency performance vs strategic plan goals (4).	Substituted GPRA planning and reporting process with program-by-program review; performance goals tied to programs vice agency strategic goals.	FGPPs must identity agencies, organizations, program activities, regulations, tax expenditures, and other activities contributing to each CAP goal. OMB must consult Congress for out of cycle changes to CAP goals. Common performance indicators with quarterly targets for CAP goals. ASPs must describe how congressional views informed agency goals and objectives. APP goals must describe how agency will achieve them. Agency Head must notify Congress when APPs are posted to agency website
OMB Oversight	Allowed to approve non-statutory waivers for managerial flexibility; charter pilot projects	Expanded OMB review of program performance; PIC created via EO 13450 (2007)	New OMB/DDM role. Chairs PIC. DDM can direct agency detailees to PIC. Designates lead USG official for CAP goals. Structures APG process and decides number of APGs government-wide and by agency. Annually reviews unmet APP goals and performance improvement plans and

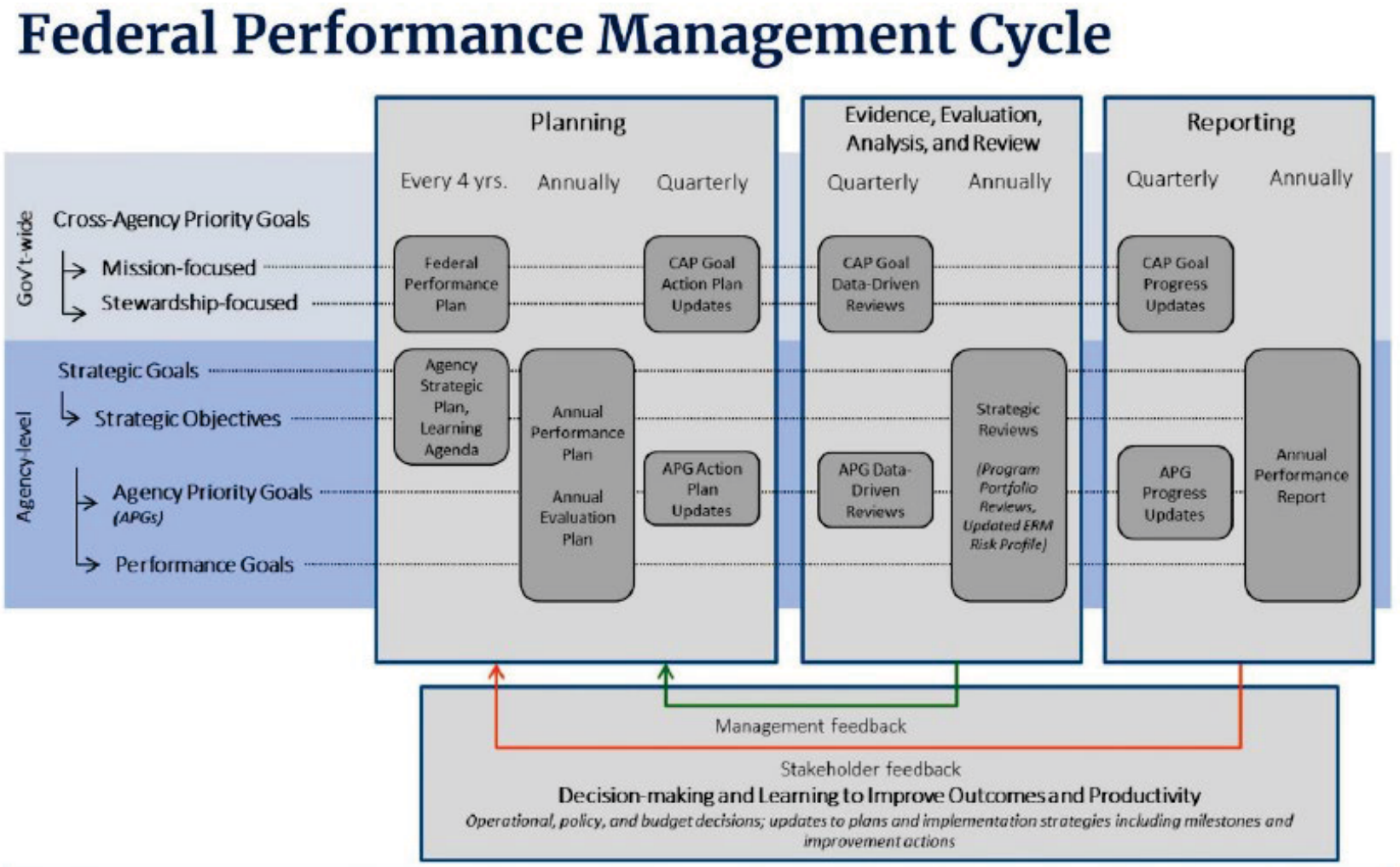
			proposed changes to budget obligations, transfers, or reprogramming to address these.
Agency Leadership	Agency heads deliver ASPs to OMB and Congress, APPs to OMB	Continued GPRA, and PIOs created via EO 13450 (2007)	Head of Agency (HOA) may adjust ASP as needed. Deputy HOAs are COOs. HOA approves APGs and with COO input appoints PIOs (5). PIOs report directly to COO. APG and Strategic Objective (6) Goal Leaders. CHCOs lead part of APP development. Lead officials to resolve identified "major management challenges." Designated senior official to lead performance improvement strategy for unmet APP goals.
Reviews of Performance	Annual *ex-post* agency-level reports to President and Congress; Performance plans must explain reasons for and plan to address unmet performance goals.	Annual evidence-based program performance evaluations to OMB.	OMB and PIC reviews progress towards CAP goal targets quarterly with Lead Official. HOA, Deputy HOA(COO), and PIO review APG progress quarterly with Goal Leaders, assess risk of not meeting APGs, identify performance improvement strategies. PIOs coordinate use of performance measures in personnel appraisals and planning and personnel processes. Annual Strategic Reviews of progress towards ASP and APP objectives (7).
Performance Data Quality	OMB must determine if program activity performance goals can be measured in an "objective, quantifiable, and measurable form."	OMB issued standardized guidance for PART scoring that required evidence and quantification.	APPs must describe how agencies will ensure performance data accuracy and reliability (i.e., how measured values are verified and validated, list of data sources, required levels of accuracy, data limitations and agency methods to compensate).
Integration with Budgeting			
Input to President's Budget	Annual performance plans informed President's Budget but agencies were not required to revise based on Congressional action.	PMA initiative for budget-performance integration. OMB evaluated agency program performance reports for President's Budget.	APPs must cover each program activity in agency budget, describe major management challenges, and designate low-priority program activities based on evidence.
Input to Congressional Budget Reviews	Directed OMB conduct at least 5 performance budgeting pilot projects.	Performance budget justifications, and *Performance and Management Assessments* as an annex, in President's Budget.	The FGPP, CAP goals, and all APPs accompany the President's Budget to Congress.
Congressional Oversight			
Plans and Goal Setting	Agencies must consult Congress when developing strategic plans	OMB reported PART evaluations with budget submission.	OMB consults "periodically" (at least biannually) with majorities and minorities of committees on CAP goals; Agencies must consult with minority and majority members

			of authorizing, appropriations, and oversight committees on ASPs. OMB reports on unmet APP goals to Congress and GAO annually.
Administrative Burdens of Reform	Not addressed in GPRA.		PIC may recommend streamlined performance management policies and requirements. Heads of agencies may consider administrative burden when deciding to hold interim quarterly performance reviews. Agencies assess plans and reports to Congress they deem as outdated or duplicative and consult with committees on product usefulness.
Transparency			
Citizen Access to Performance Information	Not addressed in GPRA beyond communications to Congress.	Agency-formatted performance information on agency websites. Program performance and scorecard ratings on public websites (results.gov and expectmore.gov)	FGPP, ASP, APP and APU reports, program information, and some CAP goal and APG information posted to OMB performance website (performance.gov). Quarterly updates. APU reports have 5-year lookback.

Sources: 1. Government Performance and Results Act of 1993, Pub. L. No. 103-62, §1, 107 Stat. 285 (1993).
2. Government Performance and Results Modernization Act of 2010, Pub. L. No. 111-352, §1, 124 Stat. 3866 (2011).
3. U.S. Library of Congress, Congressional Research Service, *The Bush Administration's Program Assessment Rating Tool (PART),* by Clinton T. Brass, RL32663 (2004).
4. U.S. Library of Congress, Congressional Research Service, *Performance Management and Budgeting in the Federal Government: Brief History and Recent Developments*, by Virginia A. McMurtry, RL32164 (2005).
5. Executive Office of the President, U.S. Office of Management and Budget, *OMB Circular No. A-11: Preparation, Submission, and Execution of the Budget* (Washington, DC, 2021), part 6.
6. Breul, Jonathan D., and John M. Kamensky. 2008. "Federal Government Reform: Lessons from Clinton's 'Reinventing Government' and Bush's 'Management Agenda' Initiatives." *Public Administration Review* 68 (6): 1009–26. https://doi.org/10.1111/j.1540-6210.2008.00950.x.

(1) Management improvement goals focus on agency financial management, human capital, IT, procurement and acquisition, and real property management.
(2) The website performance.gov states that the website in its entirely represents the federal government performance plan.
(3) Agencies may submit PARs if they exercise this option under the Reports Consolidation Act of 2000 (Pub. L. 106-531) to integrate its performance update with its financial statements.
(4) GPRA required agencies begin reporting preceding fiscal year (FY) performance beginning in FY 2000.
(5) The Congressional Research Service noted GPRAMA does not require the primary duty of officials appointed as PIOs to be their primary duty, and they may be political appointees or career officials.
(6) Goal leaders for Strategic Objectives in agency strategic plans and Deputy Goal Leaders were directed by OMB memorandum subsequent to GPRAMA signature. (U.S. Government Accountability Office 2015, 7).
(7) Annual agency strategic reviews were mandated by OMB memorandum in 2012 (U.S. Government Accountability Office 2015, 8).

Appendix B. The Federal Performance Management Cycle (as of March 2021)



Source: Executive Office of the President, U.S. Office of Management and Budget, *OMB Circular No. A-11: Preparation, Submission, and Execution of the Budget* (Washington, DC, 2021), part 6, page 6 of Executive Summary, https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf.

Appendix C. Variable Definitions, Survey Questions, and Descriptive Statistics from 2017 GAO Survey of Federal Managers

Dependent Variable	Definition and Related Survey Questions	N	Mean	SD	AIC	Cronbach's Alpha	Eigenvalue	Factor Loading
Adoption of Performance Measures (PM)	Respondent perceptions on the degree to which the organization has adopted performance measures. A key indicator of organizational use of a performance management system (factored survey questions include all five below).	2,151	3.916	1.023	.540	.854	3.035	
	To what extent, if at all, do you agree with the following statements for the program(s)/operations(s)/project(s) that you are involved with? (select one answer in each row)							
Output Measures	We have performance measures that tell us how many things we produce or services we provide.	2,333	3.776	1.097				.672
Efficiency Measures	We have performance measures that tell us if we are operating efficiently.	2,344	3.422	1.132				.797
Customer Satisfaction Measures	We have performance measures that tell us whether or not we are satisfying our customers.	2,296	3.331	1.189				.774
Quality Measures	We have performance measures that tell us about the quality of the products or services we provide.	2,336	3.425	1.154				.845
Outcome Measures	We have performance measures that would demonstrate to someone outside of our agency whether or not we are achieving our intended results.	2,342	3.672	1.053				.796
Use of Performance Information (PI)	Respondent perceptions on the degree to which organization managers at all levels use performance information to make management decisions. A second key indicator of organizational use of a performance management system (factored survey questions include all 9 below).	1,643	3.760	1.082	.686	.952	6.836	
	For those program(s)/operation(s)/project(s) that you are involved with, to what extent, if at all, do you use the information obtained from performance measurement when participating in the following activities?							
Program Strategy	Developing program strategy.	2,535	3.406	1.132				.918
Program Priorities	Setting program priorities.	2,543	3.525	1.113				.919

Resource Allocation	Allocating resources.	2,507	3.424	1.142				.881
Problem Identification	Identifying program problems to be addressed.	2,564	3.584	1.085				.926
Corrective Actions	Taking corrective action to solve program problems.	2,568	3.540	1.091				.918
Work Processes	Adopting new program approaches or changing work processes.	2,546	3.377	1.118				.908
Performance Goals	Setting new or revising existing performance goals.	2,517	3.313	1.154				.856
Contract Management	Developing and managing contracts.	1,846	2.953	1.293				.668
Program Consolidation	Streamlining programs to reduce duplicative activities.	2,363	2.962	1.188				.818

Independent Variable	**Definition and Related Survey Questions**	**N**	**Mean**	**SD**	**AIC**	**Cronbach's Alpha**	**Eigenvalue**	**Factor Loading**
External Oversight (O)	The degree to which managers perceive there is oversight of the organization's performance by external entities (factored survey questions include all five below).	1,275	3.097	1.131	.635	.897	3.526	
	To what extent, if at all, do you believe that the following persons or entities pay attention to your agency's use of performance information in management decision making?							
	Office of Management and Budget (OMB).	1,760	3.09	1.80				.879
	Federal agencies other than OMB	1,501	2.48	1.58				.868
	Congressional committees (e.g., members, staff).	1,697	2.83	1.71				.898
	The audit community (e.g. GAO, Inspectors General).	1,811	3.21	1.82				.868
	The general public.	1.813	2.68	1.50				.663
Goal Clarity (G)	The degree to which managers perceive the strategic goals of the organization are clear and measurable (factored survey questions include all four below).	1,607	2.488	1.150	.606	.860	2.782	
	To what extent, if at all, do you agree with the following statements as they relate to agency priority goals?							
	My agency's priority goals reflect the highest priorities of my agency.	1,892	3.728	1.040				.858

	My agency priority goals measure if the program is achieving its intended purpose, instead of focusing solely on what can easily be measured.	1,679	3.487	1.073			.861
	My agency has communicated to employees about its progress toward achieving priority goals.	1,944	3.340	1.165			.832
	The program(s)/operation(s)/ project(s) I am involved with contribute to the achievement of one or more of my agency's priority goals.	1,948	3.643	1.170			.716
Culture of Accountability and Empowerment (C)	The degree to which managers perceive the organizational climate fosters employee empowerment, accountability, recognition, a results orientation, and collaboration (factored survey questions include all six below),	2,386	4.247	.934	.489	.852	3.345
	To what extent, if at all, do you agree with the following statements?						
	Agency managers/supervisors at my level have the decision making authority they need to help the agency accomplish its strategic goals.	2,664	3.239	1.072			.657
	Agency managers/supervisors at my level are held accountable for agency accomplishment of its strategic goals.	2,644	3.695	.988			.832
	Agency managers/supervisors at my level are held accountable for the results of the program(s)/operation(s)/project(s) they are responsible for.	2,656	3.872	.949			.826
	Employees in my agency receive positive recognition for helping the agency accomplish its strategic goals.	2,660	3.410	1.064			.727
	My agency's top leadership demonstrates a strong commitment to achieving results.	2,612	3.767	1.096			.735
	Agency managers/supervisors at my level use performance information to share effective program approaches with others.	2,504	3.213	1.092			.685

Credible Leadership (L)	The degree to which managers perceive the senior leadership of the organization is committed to using performance management for decision making (factored survey questions include all four below).	1.464	3.547	1.119	.592	.853	2.674	
	To what extent, if at all, do you agree with the following statements?							
	My agency's top leadership demonstrates a strong commitment to using performance information to guide decision making.	2,399	3.423	1.093				.766
	Agency managers/supervisors at my level take steps to align program performance measures with agency-wide goals and objectives.	2,572	3.586	1.012				.676
	To what extent, if at all, you believe that the following person or entities pay attention to your agency's use of performance information in decision making?							
	Department Secretary or Head of Independent Agency.	1,983	3.309	1.209				.898
	Head of Major Component of Cabinet Level Department (if applicable).	1,612	3.174	1.251				.907
Training in Performance Management (T)	The degree to which managers perceive training would help employees perform a range of performance management tasks (factored survey questions include all six below).	2,604	1.557	.450	.668	.923	5.246	
	During the past 3 years, has your agency provided, arranged, or paid for training that would help you to accomplish the following tasks? (1=No, 2=Yes) [reversed]							
	Conduct strategic planning.	2,699	1.513	.500				.862
	Set program performance goals.	2,693	1.543	.498				.956
	Develop program performance measures.	2,688	1.499	.500				.959
	Assess the quality of performance data.	2,682	1.432	.500				.948
	Use program performance information to make decisions**.**	2,672	1.481	.500				.955
	Link the performance of program(s)/operation(s)/project(s) to the achievement of agency strategic goals.	2,675	1.517	.500				.928

Capacity for Performance Evaluation (X)	The degree to which managers perceive the organization has the capacity, in terms of staff, tools, and resource investments in quality and capacity, to evaluate its performance (factored survey questions include all four below).	2,021	3.388	1.008	.579	.846	2.496	
	To what extent, if at all, you agree with the following statements?							
	My agency has sufficient analytical tools for managers at my level to collect, analyze, and use performance information	2,549	3.126	1.127				.779
	The program(s)/operation(s)/project(s) that I am involved with have sufficient staff with the knowledge and skills needed to analyze performance information.	2,579	3.101	1.105				.716
	My agency is investing the resources needed to ensure that its performance data is of sufficient quality.	2,388	3.059	1.102				.856
	My agency is investing in resources to improve the agency's capacity to use performance information.	2,182	3.127	1.109				.802
Maturity of Performance Measures (M)	The degree to which managers perceive the performance measurement system in the organization is mature based on metrics development and standardization and data availability and timeliness (factored survey questions include all four below)	2,338	3.059	1.102	.587	.851	2.547	
	Based on your experience with the program(s)/operation(s)/project(s) that you are involved with, to what extent, if at all, have the following factors hindered measuring performance? [reversed]							
	Difficulty determining meaningful measures.	2,551	3.034	1.150				.720
	Different parties are using different definitions to measure performance.	2,440	3.137	1.179				.715
	Difficulty obtaining valid or reliable data.	2,504	3.271	1.173				.890
	Difficulty obtaining data in time to be useful.	2,486	3.344	1.164				.852

Notes:

All items load strongly with sufficient inter-item covariance and reliability (alpha).

No rotations due to single factor desired.
N = Number of observations
Mean = Arithmetic average
SD = Standard Deviation
AIC = Average inter-item correlation
Cronbach's Alpha = Measure of reliability
Eigenvalue = How much of observed variable's variance is explained by the factors

Appendix D. Descriptive Information on 2017 GAO Survey Questions for Factor Analysis

Question	N	Min.	Max.	Mean	Std. Dev.
5a	2,333	1	5	3.77625	1.09678
5b	2,344	1	5	3.42236	1.13228
5c	2,296	1	5	3.33101	1.18899
5d	2,336	1	5	3.42551	1.1539
5e	2,342	1	5	3.67208	1.05349
7a	2,535	1	5	3.40631	1.13207
7b	2,543	1	5	3.52536	1.11356
7c	2,507	1	5	3.42361	1.14198
7d	2,564	1	5	3.58424	1.08564
7e	2,568	1	5	3.5405	1.09106
7f	2,546	1	5	3.37667	1.11813
7i	2,517	1	5	3.31307	1.15447
7l	1,846	1	5	2.95341	1.29288
7o	2,363	1	5	2.96234	1.18765
11e	1,760	1	5	3.1858	1.20409
11f	1,501	1	5	2.66556	1.16393
11g	1,697	1	5	2.89157	1.23816
11h	1,811	1	5	3.2794	1.18709
11i	1,813	1	5	2.48759	1.15014
17b	1,892	1	5	3.7278	1.04038
17c	1,679	1	5	3.4872	1.07343
17d	1,944	1	5	3.34002	1.16481
17e	1,948	1	5	3.64271	1.17046
9a	2,664	1	5	3.23874	1.07198
9b	2,644	1	5	3.69516	0.98845
9c	2,656	1	5	3.87199	0.94927
9e	2,660	1	5	3.41015	1.06458
9g	2,612	1	5	3.76608	1.0963
9m	2,504	1	5	3.21286	1.09251
9h	2,399	1	5	3.42268	1.0935
9j	2,572	1	5	3.58631	1.01215
11a	1,983	1	5	3.30862	1.20902
11b	1,612	1	5	3.17432	1.2507
13a	2,698	0	1	0.51371	0.4999
13b	2,692	0	1	0.54309	0.49823
13c	2,687	0	1	0.4987	0.50009
13d	2,681	0	1	0.4323	0.49549

13e	2,671	0	1	0.48072	0.49972
13f	2,674	0	1	0.51683	0.49981
6e	2,549	1	5	3.12632	1.12683
6h	2,579	1	5	3.10081	1.10535
9f	2,388	1	5	3.05863	1.10257
9i	2,182	1	5	3.12741	1.10863
8a	2,551	1	5	3.0345	1.15044
8b	2,440	1	5	3.1373	1.17877
8c	2,504	1	5	3.27117	1.17282
8d	2,486	1	5	3.34433	1.16406

Appendix E. Descriptive Information on Model Variables by Federal Agency

Agency	PM	PI	O	G	C	L	T	X	M
Agriculture									
N	114	78	67	69	108	78	116	95	110
Mean	-0.18	-0.04	-0.01	-0.19	0.00	0.04	0.13	-0.06	-0.12
USAID									
N	71	65	58	64	75	52	88	75	75
Mean	0.13	0.30	0.33	0.08	0.03	0.20	0.12	0.16	-0.15
Commerce									
N	102	82	68	68	107	75	118	91	107
Mean	0.08	0.08	-0.03	-0.07	0.00	-0.04	-0.05	0.14	0.12
Defense									
N	106	93	60	82	117	79	128	109	118
Mean	0.01	0.17	-0.23	-0.33	0.05	-0.10	0.09	-0.07	0.04
Justice									
N	36	32	20	33	48	25	53	38	47
Mean	0.05	-0.06	0.27	0.15	0.06	0.00	0.18	0.39	0.11
Labor									
N	99	62	52	73	111	81	118	100	104
Mean	0.10	-0.03	0.01	-0.16	0.22	0.41	-0.13	-0.04	0.21
Energy									
N	99	82	66	71	115	71	119	88	112
Mean	0.00	0.12	-0.13	-0.09	-0.08	-0.14	0.06	-0.04	0.11
Education									
N	88	49	51	61	90	68	99	80	89
Mean	0.04	-0.10	-0.09	-0.24	-0.23	-0.13	0.04	-0.19	-0.06
EPA									
N	98	88	54	91	112	56	127	93	115
Mean	-0.21	0.01	0.08	0.14	0.01	0.17	-0.36	-0.02	0.01
GSA									
N	106	89	54	74	106	53	107	94	100
Mean	0.33	0.19	0.13	0.05	0.25	0.28	0.04	0.34	0.17
HHS									
N	104	86	65	68	120	82	140	105	123
Mean	-0.22	0.05	-0.29	-0.27	-0.17	-0.21	0.06	-0.12	-0.14
DHS									
N	86	83	67	66	113	78	121	94	108
Mean	-0.39	-0.45	-0.15	-0.19	-0.20	-0.26	-0.17	-0.16	-0.25
HUD									
N	95	63	47	79	89	68	102	79	95
Mean	-0.05	-0.02	-0.18	-0.02	-0.09	-0.09	0.22	-0.18	-0.01
Interior									

N	102	98	57	86	115	66	126	95	114
Mean	-0.17	-0.08	-0.06	-0.19	-0.01	-0.41	-0.11	-0.05	-0.08
NSF									
N	44	32	26	41	51	18	58	47	49
Mean	0.34	0.30	0.52	0.32	0.25	0.51	0.24	0.50	-0.01
NASA									
N	81	79	38	78	103	42	109	66	97
Mean	0.38	0.31	0.06	0.64	0.28	0.06	0.11	0.20	0.03
NRC									
N	84	71	41	0	94	32	100	77	90
Mean	-0.16	-0.08	0.28		0.18	0.17	-0.31	-0.05	0.15
OPM									
N	74	40	44	63	73	42	85	67	81
Mean	0.23	0.08	0.28	-0.06	0.26	0.18	0.03	-0.15	-0.14
SBA									
N	80	50	51	67	82	52	88	73	85
Mean	-0.14	-0.11	0.13	0.12	0.12	0.10	-0.13	-0.19	0.02
State	`								
N	93	69	50	56	99	70	118	89	107
Mean	-0.25	-0.19	0.14	-0.57	-0.32	-0.20	0.08	-0.20	-0.19
SSA									
N	116	71	68	109	130	77	138	112	125
Mean	0.20	0.03	0.26	0.33	0.06	0.30	0.06	0.23	0.08
Transportation									
N	96	61	65	79	102	70	115	87	96
Mean	-0.17	-0.30	-0.37	0.06	-0.19	-0.32	0.01	-0.15	-0.07
Treasury									
N	87	55	50	54	99	63	121	83	95
Mean	0.18	-0.14	-0.12	0.00	-0.10	0.12	-0.09	-0.01	0.19
VA									
N	90	65	56	75	94	66	109	84	96
Mean	0.15	-0.06	-0.13	0.31	-0.05	0.10	0.12	0.12	0.02

Notes:
PM Adoption of Performance Measures
PI Use of Performance Information
O External Oversight
G Goal Clarity
C Culture of Accountability and Empowerment
L Credible Leadership Commitment
T Performance Management Training
X Capacity for Performance Evaluation
M Measurement System Maturity

All factor variables are standardized, so mean is the number of standard deviations from universal mean of all CFI Act agencies
Standard deviations not shown since means are already standardized.

Appendix F. Fixed Effects of Agency Identity on Agency Adoption of Five Kinds of Performance Measures

	Performance Measure Adoption				
Bureaucracy Type/Agency	**Outputs**	**Efficiency**	**Customer Service**	**Quality**	**Outcomes**
Production					
Commerce		2.065 (.491)**	1.934 (.547)*	1.921 (.484)**	
GSA		3.507 (.888)**	3.663 (.893)**	2.249 (.558)**	1.565 (.359)+
Labor		2.942 (.713)**			
NRC		1.940 (.459)**			
NSF		4.487 (1.403)**	2.594 (.766)**	2.409 (.732)**	
SBA	2.435 (.661)**	1.745 (.479)*			
SSA	1.808 (.394)**	3.069 (.705)**	2.159 (.553)**		
USAID	1.211 (.340)+	1.926 (.507)*		2.146 (.582)**	
VA		2.506 (.657)**	2.499 (.628)**	2.095 (.542)**	
Procedural					
Education		2.695 (.673)**	1.854 (.496)*		
HUD		2.539 (.572)**			
OPM		4.067 (1.159)**	1.949 (.582)*	2.269 (.661)**	
Craft					
Energy		1.822 (.406)**	1.798 (.451)*		
EPA					
NASA		2.665 (.674)**		2.956 (.855)**	2.054 (.586)*
Coping					

Agriculture		1.997 (.490)**			.705 (.140)+
Defense		1.960 (.440)**	1.954 (.473)**	1.694 (.417)*	
DHS		2.015 (.529)**			.534 (.133)*
HHS					
Interior					.696 (.136)+
Justice		2.289 (.594)**			
State	.626 (.134)*				.557 (.137)*
Transportation		1.784 (.421)*			
Treasury		2.989 (.709)**	2.571 (.674)**	2.255 (.577)**	

Notes:
The five type indicators of performance measure adoption in this table are estimated for each agency using model 4.
Fixed effect generalized ordinal logits.
Values are odds ratio with EPA as the base case.
** p<.01, * p<.05., + p<.10
Robust standard errors in parentheses.

Appendix G. Fixed Effects of Agency Identity on Uses of Performance Information

	Manager Use of Performance Information								
Bureaucracy Type/Agency	**Program Strategy**	**Program Priorities**	**Allocate Resources**	**Identify Problems**	**Corrective Actions**	**Work Processes**	**Set Goals**	**Manage Contracts**	**Consolidate Programs**
Production									
Commerce									
GSA					1.604 (.365)*	1.575 (.374)+		1.567 (.337)*	1.500 (.340)+
Labor									
NRC	.558 (.128)*						.671 (.155)+		
NSF				2.513 (.891)**	2.051 (.540)**	2.050 (.667)*	1.689 (.485)+		2.141 (.596)**
SBA					.656 (.167)+			.606 (.178)+	
SSA				1.496 (.328)+	1.483 (.345)+				
USAID	1.758 (.464)*							2.363 (.610)**	1.605 (.410)+
VA									
Procedural									
Education									
HUD							.614 (.151)*		

OPM									
Craft									
Energy								1.513 (.378)+	
EPA									
NASA	.637 (.160)+			1.597 (.386)+	1.498 (.360)+			2.055 (.533)**	2.005 (.487)**
Coping									
Agriculture									
Defense									
DHS	.452 (.108)**	.433 (.099)**	.526 (.131)**	.638 (.147)+	.547 (.126)**	.630 (.152)+	.481 (.104)**	.450 (.112)**	.581 (.129)*
HHS									
Interior									
Justice						.593 (.181)+			
State	.646 (.130)*	.598 (.124)*		.685 (.142)+	.582 (.122)**	.619 (.140)*	.649 (.130)*		
Transportation	.490 (.113)**	.576 (.132)*	.656 (.154)+			.527 (.125)**	.621 (.122)*	.578 (.150)*	.684 (.154)+
Treasury		.687 (.148)+						.519 (.158)*	

Notes: EPA is the base case
The nine indicators of performance information use in this table are estimated for each agency using model 4.

Fixed effect generalized ordinal logits.
Values are odds ratio with EPA as the base case.
** p<.01, * p<.05., + p<.10
Robust standard errors in parentheses.

Appendix H. Survey for Federal Performance Managers in CFO Act Agencies (2021)

GPRAMA Institutionalization Survey

Please enter the personal access code that was in the registration email.

_____

Welcome! Before we proceed, please read the following information regarding informed consent.

Principal Investigator: Matthew Dull, Ph.D; Co-Investigator: Edgar Hollandsworth, Ph.D. candidate.

IRB# and Title of Study: 20-814 Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)

Sponsor: Not applicable.

You are invited to participate in a research study. This section includes information about the study and contact information if you have any questions.

I am a graduate student at Virginia Tech, and I am conducting this research as part of my course work.

WHAT SHOULD I KNOW? If you decide to participate in this study, you will be asked to complete an online survey and/or an online interview, depending on your employment situation. As part of the study, you will be asked to take an online survey of about 15 minutes in duration if you are a federal manager. If you are employed by the Social Security Administration, the Small Business Administration, National Aeronautics and Space Administration, the Department of Energy, the Department of Homeland Security, Department of Treasury, Department of Education, or the Department of Housing and Urban Development, you will also be invited to meet with the researcher for about one hour in a recorded online videoconference for a structured interview to clarify any questions and responses from the online survey and to answer several additional questions. If you are a fellow of the National Academy for Public Administration (NAPA), you will be invited to participate in a 1-hour recorded online interview only and should not take this survey.

The study should take approximately 15 to 75 minutes of your time.

We do not anticipate any risks from completing this study. Because the research will take place only online, it is not expected your participation will expose you to any risks which are different from those you would experience in a daily work routine. Questions in the survey and interviews will ask you, as a government employee, to reflect on situations and processes in your workplace for which you may be responsible. Describing management challenges in your organization may be stressful. To alleviate any such concerns, your responses will not be attributed to you in study reports or presentations. In addition, copies of your survey responses and interview transcripts will be provided to you on request, but shared with no one else. No

classified information will be requested or discussed. You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

CONFIDENTIALITY. We will do our best to protect the confidentiality of the information we gather from you, but we cannot guarantee 100% confidentiality. Any data collected during this research study will be kept confidential by the researchers. Your interview will be audio-recorded using a digital recorder and then transcribed. The researchers will code the transcripts using a pseudonym (false name). The recordings will be uploaded to a secure cloud storage space using strong AES256 encryption. The researchers will maintain a list that includes a key to the code. The master key and the recordings will be stored for 2 years after the study has been completed and then destroyed.

WHO CAN I TALK TO? If you have any questions or concerns about the research, please feel free to contact Edgar Hollandsworth at hoya86@vt.edu. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research participant, contact the Virginia Tech HRPP Office at 540-231-3732 (irb@vt.edu).

There are 8 sets of closed-ended questions, each with several items. There is also an option for "other" for each question for free-form responses. The total time required should be 15 minutes or less. Thank you for participating.

Q1. To report organizational performance, organizations use a variety of performance measures. Based on your observations, please select "yes" if your organization routinely uses the measure when reporting its performance in support of GPRAMA, and "no" if it does not.

	Yes	No	Don't know
Customer service measures assess service delivery to a customer, client, citizen, or other recipient, which can include an assessment of quality, timeliness, and satisfaction among other factors.	○	○	○
Efficiency measures are the ratio of a program activity's outputs (amount of products or services delivered) to its inputs (such as costs or hours worked by employees).	○	○	○
Cost-effectiveness measures are the ratio of outcomes (results of program activity) to inputs (such as costs or hours worked by employees.	○	○	○
Output measures tabulate, calculate, or record activity or effort that can be expressed in a quantitative or qualitative manner.	○	○	○
Outcome measures assess the results of a program activity compared to its intended purpose.	○	○	○
Alternative measures approved by the Office of Management and Budget are generally not objective, quantifiable, and measurable.	○	○	○
Other (please specify):	○	○	○

Q2. Effective analysis of performance data depends on different conditions and the use of certain evaluation practices in an organization. Based on your observations, please select "yes" if the condition or practice is routine in your organization when evaluating its performance in support of GPRAMA, and "no" if it is not.

	Yes	No	Don't know
Performance analysts understand the cause and effect relationships when measuring performance.	○	○	○
Performance analysts understand what is being measured using performance data.	○	○	○
The results of performance analysis accurately reflect managerial achievements.	○	○	○
Baseline information for comparisons of performance over time is available.	○	○	○
Most forms of performance analyzed are measurable and quantifiable.	○	○	○
Data needed for performance analysis are available.	○	○	○
Data needed for performance evaluation are factual and unbiased.	○	○	○
Data needed for performance analysis are timely.	○	○	○

Q3. Executives and managers at many levels can use performance information for a variety of decisions. Based on your observations, please select "yes" if the practice is routine in your organization, and "no" if it is not.

	Yes	No	Don't know
Managers use PI to adopt new program approaches or change work processes.	○	○	○
Managers use PI to streamline programs to reduce duplicative activities.	○	○	○
Managers use PI to take corrective action to solve program problems.	○	○	○
Managers use PI to set new or revise existing performance goals.	○	○	○
Managers use PI to allocate or to re-allocate resources.	○	○	○
Managers use PI to develop and manage contracts.	○	○	○
Managers use PI to develop program strategies.	○	○	○
Managers use PI to set program priorities.	○	○	○
Other (please specify):	○	○	○

Q4. Please check the appropriate circle to indicate your level of agreement with the following statements based on your observations or based on statements by executive-level leaders in your organization. *In this survey, "political leaders" are elected officials or their staffs, or political appointees. "Executive leaders" are employees who are members of the Senior Executive Service or its equivalent in your organization. "Managers" are supervisors at the GS-13, GS-14, or GS-15 level, or their equivalent in your organization.*

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Don't know
Political leaders are influenced by interest groups to advocate for use of the groups' desired measures of performance.	○	○	○	○	○	○
Political leaders are influenced by interest groups to interpret performance information to suit the groups' agendas.	○	○	○	○	○	○
Political leaders use performance information to hold executives or managers accountable for program outcomes--regardless of their level of responsibility for them.	○	○	○	○	○	○
Political leaders want the agency to use the same performance management systems as other agencies.	○	○	○	○	○	○
Political leaders expect the agency to improve its performance to match other agencies'.	○	○	○	○	○	○
Political leaders want the agency to conform to published performance management standards.	○	○	○	○	○	○
Executive leaders in my organization think reporting performance is required to justify budget requests.	○	○	○	○	○	○
Executive leaders in my organization respond to conflicting stakeholder interests and policy priorities.	○	○	○	○	○	○
Executive leaders in my organization rely on partnerships and networking to manage.	○	○	○	○	○	○

Other (please specify):	○	○	○	○	○	○

Q5. Please check the appropriate circle to indicate your level of agreement with the following statements based on your observations about your organization.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Don't know
Managers use performance information to collaborate with others on effective approaches.	○	○	○	○	○	○
Managers are empowered with the authority they need to accomplish strategic goals.	○	○	○	○	○	○
Managers are held accountable for agency strategic goal accomplishment.	○	○	○	○	○	○
Managers are held accountable for goals they are responsible for.	○	○	○	○	○	○
Executive leaders demonstrate a commitment to achieving results.	○	○	○	○	○	○
Employees receive recognition for goal accomplishment.	○	○	○	○	○	○
Other (please specify):	○	○	○	○	○	○

Q6. Please check the appropriate circle to indicate your level of agreement with the following statements based on your observations about your organization. Each statement below completes this sentence: **"My agency's executive-level leadership demonstrates a strong commitment to using performance management in my organization by. . ."**

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Don't know
. . .aligning program performance measures with strategic plan goals and objectives.	○	○	○	○	○	○
. . .fostering innovations that improve performance management capabilities.	○	○	○	○	○	○
. . .using performance information to guide routine decision making.	○	○	○	○	○	○
. . .holding themselves accountable for organizational performance.	○	○	○	○	○	○
. . .publicizing organizational performance management initiatives.	○	○	○	○	○	○
. . .considering performance information when setting budgets.	○	○	○	○	○	○
. . .devoting time and attention to performance management.	○	○	○	○	○	○
Other (please specify):	○	○	○	○	○	○

Q7. Please check the appropriate circle to indicate your level of agreement with the following statements based on your observations about your organization.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Don't know
The organization has enough staff with the knowledge and skills to analyze performance information.	○	○	○	○	○	○
The organization has learning forums where employees develop skills in performance management.	○	○	○	○	○	○
The organization has analytical tools and methods to collect, analyze, and use performance information.	○	○	○	○	○	○
The organization is investing resources to improve its capacity to use performance information.	○	○	○	○	○	○
The organization is investing resources to generate quality performance data.	○	○	○	○	○	○
Other (please specify):	○	○	○	○	○	○

Q8. Please check the appropriate circle to indicate your level of agreement with the following statements based on your observations about your organization.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Don't know
The performance measurement system links performance information to organizational goals and strategic plans.	○	○	○	○	○	○
Managers are able to obtain performance information in time to be useful for decision making.	○	○	○	○	○	○
Employees use the same definitions to measure organizational performance.	○	○	○	○	○	○
Performance reports make use of benchmarks for performance measures.	○	○	○	○	○	○
Managers can get performance data they assess to be valid and reliable.	○	○	○	○	○	○
The performance measurement system uses different kinds of data.	○	○	○	○	○	○
Performance reports satisfy GPRAMA reporting requirements.	○	○	○	○	○	○
Performance goals are achievable.	○	○	○	○	○	○
Other (please specify):	○	○	○	○	○	○

Thank you for taking this survey. You have completed the survey. Your responses have been recorded and will be sent to you on request.

Appendix I. Recruiting Flier for GPRAMA Institutionalization Survey

VIRGINIA TECH

WHAT: Participants are needed for a Virginia Tech Center for Public Administration and Policy research study (IRB # 20-814).

WHY: The purpose of this research is to develop a greater understanding of factors that influence how federal agencies and senior managers implement the requirements of the Government Performance and Results Act/Modernization Act of 2010 (GPRAMA), an important government reform.

WHO: You may qualify for this research if you are a member of the federal Senior Executive Service or a federal manager at the GS-15/14/13 level (or equivalent) responsible for administering GPRAMA performance management and evaluation requirements in your organization. Also, fellows of the National Academy of Public Administration (NAPA) with previous experience in GPRAMA and its implementation in federal agencies are welcome.

HOW: Eligible federal executive and manager participants will complete a 15-minute online survey. In addition, participants from SSA, SBA, NASA, Energy, DHS, Treasury, Education, and HUD will be invited to participate in an online interview for about one hour. Selected NAPA fellows will be interviewed online as well. Online interviews will be recorded and transcribed.

FOR MORE INFORMATION: Contact Edgar Hollandsworth at hoya86@vt.edu or (703) 939-7540. Details about the study can be viewed at https://hoya86.wixsite.com/research

OTHER: No compensation will be provided, but your participation may benefit others by enabling the discovery of new management strategies that may foster agency adoption of improved performance management practices which improve the use of performance information for key management decisions. Over time, these practices can increase agency performance and citizen satisfaction with agency products or services.

Appendix J. Email Templates for Communication with Study Volunteers

Email 1 – To Federal Performance Improvement Officers [PIO] at Federal Department and Agencies

Subject: Request for Performance Improvement Officers: Virginia Tech Research Study on GPRAMA Implementation

[Date]

Dear [PIO Name],

As the Performance Improvement Officer for your department or agency, a vital part of your role is to help agency program managers and goal leaders promote the adoption of effective practices to improve agency outcomes. As a PhD candidate at Virginia Tech's Center for Public Administration and Policy in Arlington, I am currently conducting research designed to further this same objective and produce findings for theory and practice of potential value for senior federal strategic planners and performance managers during the spring of 2021.

I am writing to you to request that you share the attached flyer with a select group of individuals in your federal organization and encourage them to participate voluntarily in my study. However, please do not pressure anyone to participate. The research design calls for enrolling your Deputy Performance Improvement Officer and at least three managers at the GG-13/14/15 level, or their equivalents, who are responsible for planning, organizing, and directing your organization's performance management activities supporting the Government Performance and Results Act/Modernization Act of 2010 (GPRAMA). More information about me and my research can be found at https://hoya86.wixsite.com/research

Participation in this study (IRB # 20-814) will entail a 15-minute online survey. Participants at selected federal departments and agencies, as listed in the flyer, will also be invited to a 1-hour recorded online interview after the survey. Responses to the survey and the interviews will not be attributed to agencies or individuals. Thank you for considering this request. If you have any questions, please contact me at hoya86@vt.edu or at (703) 939-7540.

Sincerely,

Edgar M. Hollandsworth
Director of National Intelligence Pat Roberts Intelligence Scholar 2020-2021
PhD Candidate
Virginia Tech

Center for Public Administration and Policy
Arlington, VA
Hoya86@vt.edu

Enclosures
Announcement of Research Study (IRB 20-418)

Email 2 - To Federal Managers Who Respond to Recruitment Flyer

Subject: Virginia Tech Research Study on GPRAMA Implementation – How to Register

[Date]

Dear [Name],

Thank you for your interest in the announcement about the Virginia Tech Center for Public Administration and Policy research study on *Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)* (IRB 20-814). This research is supporting my dissertation. More information about the study and myself can be found at the following website: https://hoya86.wixsite.com/research.

To register as a study participant, please read the enclosed information about informed consent and then, if you consent to participate, send me an email at hoya86@vt.edu. In the email, please 1) indicate that you have read and understood the consent information, 2) confirm that you give your consent to participate, and 3) provide your name, organization, position title, and government grade or rank to validate your eligibility to participate in this study. Feel free to let me know if you have any questions at any time.

Once I receive the requested consent and information, I will email you a link to the online survey, a password, and an individual personal access code. If you are employed by the Social Security Administration, the Small Business Administration, NASA, the Dept of Energy, the Dept of Homeland Security, the Dept of Treasury, the Dept of Education or the Dept of Housing and Urban Development, I will follow up with you after you finish the survey by sending you another email to request an appointment for a 1-hour online interview using Skype or other teleconferencing system within the next two weeks.

Again, thank you for your assistance with this study. I look forward to working with you!

Sincerely,

Edgar M. Hollandsworth
Director of National Intelligence Pat Roberts Intelligence Scholar 2020-2021
PhD Candidate
Virginia Tech
Center for Public Administration and Policy

Arlington, VA
Hoya86@vt.edu

Enclosures

1. *Information Sheet for Participation in a Research Study*

Email 3 - To Federal Managers who Provide their Consent and Information

Subject: Virginia Tech GPRAMA Implementation Study – Confirming Participation

[Date]

Dear [Name],

Thank you for sending me your consent and information to participate in the Virginia Tech Center for Public Administration and Policy research study *Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)* (IRB 20-814).

In the next 24 hours, you should receive an email from the Qualtrics system at Virginia Tech that includes a link to the online survey as well as a password and a personal access code you will need to take the survey. The email's subject line will be "Virginia Tech GPRAMA Survey." Please be sure to check both your inbox and your spam folder for this email, as it is possible your email settings may route the email from Qualtrics into your spam folder instead of your inbox. If for some reason you cannot find the email, please let me know and I will re-send it.

Sincerely,

Edgar M. Hollandsworth
Director of National Intelligence Pat Roberts Intelligence Scholar 2020-2021
PhD Candidate
Virginia Tech
Center for Public Administration and Policy
Arlington, VA
Hoya86@vt.edu

Email 4 - To Federal Managers Inviting them to Interview after they Complete the Online Survey

Subject: Virginia Tech GPRAMA Implementation Study – Scheduling Interviews

[Date]

Dear [Name],

Thank you for completing the online survey for the Virginia Tech Center for Public Administration and Policy study *Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)* (IRB 20-814).

As the final step in this research, I invite you to schedule a 1-hour appointment with me to conduct a structured interview. This interview will go into greater depth on some of the survey questions, and cover some other topics that did not appear in the survey. Due to the ongoing COVID pandemic, I am scheduling surveys online using Zoom. I am flexible and will try to use a different system, if that is your preference.

All interviews will be recorded and transcribed, and your transcribed interview will be available to you on request. Note that, as with your survey responses, all interview recordings and transcriptions will be saved to a secure, encrypted folder.

Please respond via email and let me know several days and times in the next two weeks that would be convenient for you to meet with me for about an hour. I respect your schedule, so the two-week window is preferred, but not hard and fast.

Sincerely,

Edgar M. Hollandsworth
Director of National Intelligence Pat Roberts Intelligence Scholar 2020-2021
PhD Candidate
Virginia Tech
Center for Public Administration and Policy
Arlington, VA
Hoya86@vt.edu

Email 5 - To NAPA Fellows Who Respond to Recruitment Flyer

Subject: Virginia Tech Research Study on GPRAMA Implementation – How to Register

[Date]

Dear [Name],

Thank you for your interest in the announcement about the Virginia Tech Center for Public Administration and Policy research study on *Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)* (IRB 20-814).

This research is supporting my dissertation. More information about the study and myself can be found at the following website: https://hoya86.wixsite.com/research.

To register as a study participant, please read the enclosed information about informed consent and then, if you consent to participate, send me an email at hoya86@vt.edu. In the email, please 1) indicate that you have read the consent information, 2) confirm that you give your consent to participate, and 3) provide your name, organization, position title, and a short description of your experience with the implementation of GPRAMA 2010.

Once I receive the requested consent and information, I will send you another email to request an appointment for a 1-hour online interview using Zoom or other teleconferencing system within the next two weeks.

Again, thank you for your assistance with this study. I look forward to working with you!

Sincerely,

Edgar M. Hollandsworth
Director of National Intelligence Pat Roberts Intelligence Scholar 2020-2021
PhD Candidate
Virginia Tech
Center for Public Administration and Policy
Arlington, VA
Hoya86@vt.edu

Enclosures

1. *Information Sheet for Participation in a Research Study*

Email 6 - To NAPA Fellows to Invite them to Interview after they Provide Requested Information

Subject: Virginia Tech Research Study on GPRAMA Implementation – Scheduling Interviews

[Date]

Dear [Name],

Thank you for sending me your information to participate in the Virginia Tech Center for Public Administration and Policy research study *Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)* (IRB 20-814).

As the final step in this research, I invite you to schedule a 1-hour appointment with me to conduct a structured interview. This interview will ask for your opinions and perspectives on a range of possible explanations for uneven patterns of acceptance and use of performance management practices in federal agencies in support of GPRAMA requirements. Due to the ongoing COVID pandemic, I am scheduling surveys exclusively online using Zoom. I am flexible and will try to use a different system, if that is your preference.

All interviews will be recorded and transcribed, and your transcribed interview will be available to you on request. Note that, as with your survey responses, all interview recordings and transcriptions will be saved to a secure, encrypted folder.

Please respond via email and let me know several days and times in the next two weeks that would be convenient for you to meet with me for about an hour. I respect your schedule, so the two-week window is preferred, but not hard and fast.

Sincerely,

Edgar M. Hollandsworth
Director of National Intelligence Pat Roberts Intelligence Scholar 2020-2021
PhD Candidate
Virginia Tech
Center for Public Administration and Policy
Arlington, VA
Hoya86@vt.edu

Appendix K. Informed Consent Information Sheet

Information Sheet for Participation in a Research Study

Principal Investigator: Matthew Dull, Ph.D; ; Co-Investigator: Edgar Hollandsworth, Ph.D. candidate.
IRB# and Title of Study: 20-814 Performance Management Reform in the U.S. Federal Government: An Analysis of Institutional and Organizational Influences on the Implementation of the Government Performance and Results Modernization Act of 2010 (GPRAMA)
Sponsor: Not applicable.

You are invited to participate in a research study. This form includes information about the study and contact information if you have any questions.

I am a graduate student at Virginia Tech, and I am conducting this research as part of my course work.

WHAT SHOULD I KNOW?

If you decide to participate in this study, you will be asked to complete an online survey and/or an online interview, depending on your employment situation. As part of the study, you will be asked to take an online survey of about 15 minutes in duration if you are a federal manager. If you are employed by the Social Security Administration, the Small Business Administration, National Aeronautics and Space Administration, the Department of Energy, the Department of Homeland Security, Department of Treasury, Department of Education, or the Department of Housing and Urban Development, you will also be invited to meet with the researcher for about one hour in a recorded online videoconference for a structured interview to clarify any questions and responses from the online survey and to answer several additional questions. If you are a fellow of the National Academy for Public Administration (NAPA), you will be invited to participate in a 1-hour recorded online interview only.

The study should take approximately 15 to 75 minutes of your time.

We do not anticipate any risks from completing this study. Because the research will take place only online, it is not expected your participation will expose you to any risks which are different from those you would experience in a daily work routine.

Questions in the survey and interviews will ask you, as a government employee, to reflect on situations and processes in your workplace for which you may be responsible. Describing management challenges in your organization may be stressful. To alleviate any such concerns, your responses will not be attributed to you in study reports or presentations. In addition, copies of your survey responses and interview transcripts will be provided to you on request, but shared with no one else. No classified information will be discussed.

You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

CONFIDENTIALITY

We will do our best to protect the confidentiality of the information we gather from you, but we cannot guarantee 100% confidentiality.

Any data collected during this research study will be kept confidential by the researchers. Your interview will be audio-recorded using a digital recorder and then transcribed. The researchers will code the transcripts using a pseudonym (false name). The recordings will be uploaded to a secure cloud storage space using strong AES256 encryption. The researchers will maintain a list that includes a key to the code. The master key and the recordings will be stored for 2 years after the study has been completed and then destroyed.

WHO CAN I TALK TO?

If you have any questions or concerns about the research, please feel free to contact **Edgar Hollandsworth at hoya86@vt.edu.** You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research participant, contact the Virginia Tech HRPP Office at 540-231-3732 (irb@vt.edu).

Please print out a copy of this information sheet for your records.

Appendix L. Summary Statistics from Survey of 20 Federal Performance Managers (2021)

Questions on Institutionalization Outcome Indicators

		N	Yes	% Yes
1	**Agency Adoption of Performance Measures**			
1.1	Customer Service Measures	20	17	85.00%
1.2	Efficiency Measures	20	13	65.00%
1.3	Cost-Effectiveness Measures	20	10	50.00%
1.4	Output Measures	20	20	100.00%
1.5	Outcome Measures	20	19	95.00%
1.6	Alternative Measures[90]	14	8	57.14%
1.7	Other Measures	2	2	100.00%
2	**Agency Uses of Performance Information**			
2.1	Adopt new program approaches or work processes	17	14	82.35%
2.2	Streamline programs to reduce duplicative activities	14	11	78.57%
2.3	Take corrective actions to solve problems	17	15	88.24%
2.4	Set new/revise performance goals	19	19	100.00%
2.5	Allocate/re-allocate resources	18	11	61.11%
2.6	Develop and manage contracts	11	5	45.45%
2.7	Develop program strategies	17	14	82.35%
2.8	Set program priorities	17	14	82.35%
2.9	Other Uses of Performance Information	1	1	100.00%

Questions on System Action and Organization Action Condition Indicators

		N	Agree	% Agree	Mean (1-5)
1	**Political Influence and Resource Dependency**				
1.1	Political leaders are influenced by interest groups to advocate for use of the groups' desired measures of performance.	14	2	14.29%	2.43
1.2	Political leaders are influenced by interest groups to interpret performance information to suit the groups' agendas.	15	4	26.67%	2.73
1.3	Political leaders use performance information to hold executives or managers accountable for program outcomes-regardless of their level of responsibility for them.	18	6	33.33%	2.89

[90] Examples included "science outcome goals" (NASA) and "input/workload" (Treasury).

1.4	Political leaders want the agency to use the same performance management systems as other agencies	17	3	17.65%	2.18
1.5	Political leaders expect the agency to improve its performance to match other agencies'.	18	6	33.33%	2.89
1.6	Political leaders want the agency to conform to published performance management standards.	19	6	31.58%	2.79
1.7	Executive leaders think reporting performance is required to justify budget requests.	19	13	68.42%	3.89
1.8	Executive leaders respond to conflicting stakeholder interests and policy priorities.	17	13	76.47%	4.18
1.9	Executive leaders rely on partnerships and networking to manage.	20	17	85.00%	4.30
1.10	Other	1	1	100.00%	4.00
2	**Accountability and Empowerment in Agency Culture**				
2.1	Managers use performance information to collaborate with others on effective approaches.	19	15	78.95%	3.68
2.2	Managers are empowered with the authority they need to accomplish strategic goals.	19	14	73.68%	3.74
2.3	Managers are held accountable for agency strategic goal accomplishment.	20	10	50.00%	3.45
2.4	Managers are held accountable for goals they are responsible for.	20	16	80.00%	3.95
2.5	Executive leaders demonstrate a commitment to achieving results.	20	17	85.00%	4.25
2.6	Employees receive recognition for goal accomplishment.	20	13	65.00%	3.70
2.7	Other	2	1	50.00%	3.50
3	**Senior Leader Credible Commitment** *My agency's executive-level leadership demonstrates a strong commitment to using performance management in my organization by. . .*				

3.1	. . .aligning program performance measures with strategic plan goals and objectives.	20	16	80.00%	4.20
3.2	. . .fostering innovations that improve performance management capabilities.	18	11	61.11%	3.67
3.3	. . .using performance information to guide routine decision making.	20	14	70.00%	3.80
3.4	. . .holding themselves accountable for organizational performance.	20	14	70.00%	3.90
3.5	. . .publicizing organizational performance management initiatives.	19	10	52.63%	3.58
3.6	. . .considering performance information when setting budgets.	20	12	60.00%	3.40
3.7	. . .devoting time and attention to performance management.	20	15	75.00%	3.75
3.8	Other	1	1	100.00%	5.00
4	**Organizational Capacity for Performance Evaluation**				
4.1	The organization has enough staff with the knowledge and skills to analyze performance information.	19	11	57.89%	3.47
4.2	The organization has learning forums where employees develop skills in performance management.	20	13	65.00%	3.60
4.3	The organization has analytical tools and methods to collect, analyze, and use performance information.	20	18	90.00%	4.15
4.4	The organization is investing resources to improve its capacity to use performance information.	20	16	80.00%	4.00
4.5	The organization is investing resources to generate quality performance data.	20	16	80.00%	4.15
4.6	Other	1	1	100.00%	5.00
5	**Maturity of Performance Measurement System**				
5.1	The performance measurement system links performance information to organizational goals and strategic plans.	20	18	90.00%	4.55
5.2	Managers are able to obtain performance information in time to be useful for decision making.	20	14	70.00%	3.75

5.3	Employees use the same definitions to measure organizational performance.	18	10	55.56%	3.50
5.4	Performance reports make use of benchmarks for performance measures.	18	16	88.89%	4.22
5.5	Managers can get performance data they assess to be valid and reliable.	19	19	100.00%	4.37
5.6	The performance measurement system uses different kinds of data.	20	18	90.00%	4.30
5.7	Performance reports satisfy GPRAMA reporting requirements.	20	20	100.00%	4.80
5.8	Performance goals are achievable.	20	18	90.00%	4.45
5.9	Other	1	1	100.00%	5.00

Agencies with participating employees:

US Agency for International Development (USAID)
Department of Commerce (DOC)
Department of Defense (DOD)
Department of Health and Human Services (HHS)
Department of Homeland Security (DHS)
Department of Housing and Urban Development (HUD)
Department of Justice (DOJ)
Department of Labor (DOL)
National Aeronautics and Space Administration (NASA)
Small Business Administration (SBA)
Department of Transportation (DOT)
Department of the Treasury (Treasury)

Appendix M. Interview Questions for Federal Agency Performance System Managers

Date:

Time:

Personal Access Code:

Introduction

Thank you for completing the online survey and agreeing to meet with me for this interview today. Before we continue, I would like to remind you that your participation in this interview is completely voluntary, you may withdraw from the study at any time, and your responses will be protected and treated as anonymous in the study findings and report. No classified information will be discussed. The complete audio file and transcribed report of this interview will be sent to you on request. This interview should take approximately one hour. Do you consent to continue with the interview?

Section 1. Organizational Structure and Processes for GPRAMA

Each federal department and agency has an internal organizational structure and routines they follow to manage its organizational and program performance and to report performance to Congress in accordance with GPRAMA requirements.

- Please describe the main leaders, groups, and business processes in your organization devoted to these tasks.
- Which internal and external stakeholders contributed to decisions about how to establish your organization's performance management system?
- How far down in the levels of the organization do managers use performance information for making decisions?
- How standardized are your organization's performance measures and business processes for analyzing performance information?
- Do you believe your organization's current level of capabilities and business processes are sufficient to fulfill all GPRAMA requirements? If not, why not? What would have to change to make this possible?

Section 2. External Politics and Organizational Dependency on Resources

In your online survey, you identified that the following kinds of external political or resource pressures exist in your organization that may affect how your organization implements GPRAMA.

[Read-back positive survey results to interviewee].

- Based on your survey responses, what have you observed about each of these influences in your agency and how you have responded to them?

- What practices have you undertaken to improve your agency's relations with political leaders that would support your agency's implementation of GPRAMA? What were the results?

Section 3. Accountability and Empowerment in Organizational Culture

In your online survey, you identified that the following kinds of cultural influences exist in your organization that may affect how your organization implements GPRAMA.

[Read-back positive survey results to interviewee].

- Based on your survey responses, what have you observed about each of these influences in your agency, and how you have responded to them?
- How has your agency tried to improve managers' sense of accountability for their responsibilities to implement GPRAMA? What were the results?

Section 4. Credible Commitment of Senior Leaders in the Organization

In your online survey, you identified that senior leaders in your organization have demonstrated a level of commitment to performance management that may affect how your organization implements GPRAMA.

[Read-back positive survey results to interviewee].

- Based on your survey responses, what have you observed about each of these influences in your agency, and how you have responded to them?
- What practices have you undertaken to demonstrate credible leadership commitment that would support your agency's implementation of GPRAMA? What were the results?

Section 5. Organizational Capacity to Evaluate Performance Data

In your online survey, you identified the following features of your organization's capacity to evaluate performance information that may affect how your organization implements GPRAMA.

[Read-back positive survey results to interviewee].

- Based on your survey responses, what have you observed about each of these influences in your agency, and how you have responded to them?
- What practices have you undertaken to improve or expand the capacity to evaluate performance that would support your agency's implementation of GPRAMA? What were the results?

Section 6. Maturity of the Organization's Performance Measurement System

In your online survey, you identified the following features of your organization's performance measurement system that may affect how your organization implements GPRAMA.

[Read-back positive survey results to interviewee].

- Based on your survey responses, what have you observed about each of these influences in your agency and how you have responded to them?
- What practices have you undertaken to improve your organization's performance measurement system that would support your agency's implementation of GPRAMA? What were the results?

Section 7. Managerial Incentives to Implement GPRAMA

1. If there were no federal law requiring federal agencies to report their performance to Congress, would you still be an advocate for measuring your organization's performance?

   [If response is "no," skip to question 4 and then ask questions 4 and 5]

2. Apart from any tangible rewards from your organization, how do you benefit professionally or personally by implementing GPRAMA in your agency?

3. Do you feel your efforts to set up performance management systems in your organization are worth your time and effort to accomplish? If not, why not?

   ------- Stop --------

4. Does your organization provide any advantages for career advancement for your efforts to implement GPRAMA?
   a. If so, do you consider these incentives to be strong or weak?

5. Does your organization provide any bonuses or other kinds of inducements for implementing GPRAMA?
   a. If so, do you consider these incentives to be strong or weak?

Section 8. Managerial Attitudes About Performance Management

Interview questions:

1. Do you believe that your organization's performance management system improves your organization's performance or reduces costs? If so, how?

2. Does GPRAMA provide your organization *instrumental* benefits (examples: internal policy changes, cultural changes, improved internal communications, leader development, advocating for a larger budget, greater autonomy)? If so, please provide examples.

3. Does GPRAMA provide your organization *symbolic* benefits (examples: advocating for external policy changes, creating positive impressions of compliance with legislation or

court orders, improving the organization's reputation, etc). If so, please provide examples.

4. Do you believe you have the amount of *authority* you need to carry out your responsibilities to implement GPRAMA? If not, what additional authorities do you need?

5. Do you believe GPRAMA requirements are in harmony with the mission of your organization? If not, why not?

Section 9: Noteworthy agency-specific characteristics affecting GPRAMA Implementation

Based on recent quantitative research using a Government Accountability Office survey of federal managers in 2017, your organization appears to have unusual strengths or opportunities for development in some of its practices for implementing GPRAMA.

[Highlight significant factors from agency fixed effects modeling]

What do you believe makes managers in your organization think that your organization is more or less effective in each of these practices?

Appendix N. Interview Questions for National Academy of Public Administration Fellows

Date:

Time:

Personal Access Code:

Introduction

Thank you for agreeing to meet with me for this interview today. Before we continue, I would like to remind you that your participation in this interview is completely voluntary, you may withdraw from the study at any time, and your responses will be protected and treated as anonymous in the study findings and report. No classified information will be discussed. The complete audio file and transcribed report of this interview will be sent to you on request. This interview should take approximately one hour. Do you consent to continue with the interview?

The Government Performance and Results Modernization Act of 2010 is now a decade old. The GAO has reported that, despite evidence many federal agencies have adopted it and have started to implement it, others have only done so partially. Repeated GAO recommendations to these agencies have often gone unanswered. As Dave Mader wrote in 2019, "despite people's best efforts and intentions, many reform efforts fall short of their goals" (NAPA 2019, 9).

Questions

1. To what do you attribute the fact that some federal departments and agencies are not moving ahead with full implementation of GPRAMA?

- What are the main challenges that agencies have identified and what are they doing to overcome them?

2. In Washington, political influence is felt everywhere, even in GPRAMA implementation. Decisions about how to measure or evaluate agency performance, in particular, may be influenced by agendas of interest groups and elected leaders.

- Have you seen any instances of politicization of GPRAMA reporting, and how did senior agency leaders respond?
- Have departmental or agency budgets ever been at any real risk for reductions due to GPRAMA reporting?

3. In your view, how important is personal engagement by top-level leadership in federal departments and separate agencies for success in implementing GPRAMA?

- Have you observed cases where internal agency resource shortfalls being an impediment to them implementing GPRAMA?
- How important is their demonstrated commitment to their agency's implementation to GPRAMA?

4. From your perspective, how effective has the Office of Management and Budget been in overseeing GPRAMA implementation?

- What challenges have they faced?
- How have agencies responded to OMB leadership and reform initiatives?

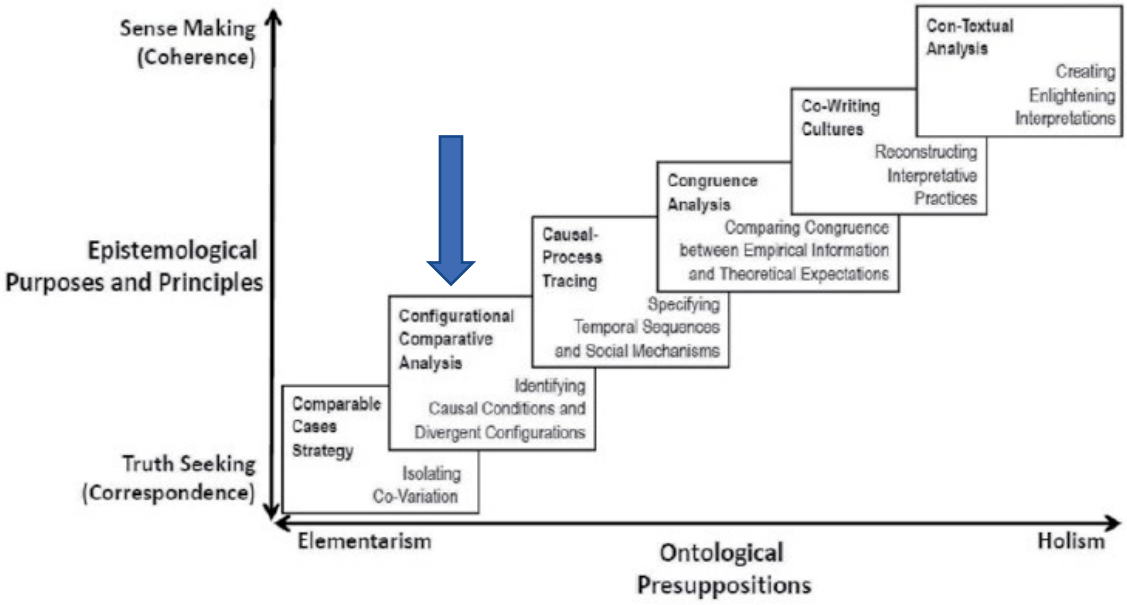
5. The President's Management Agenda in 2018 identified better use of data as one of three primary drivers of transformational change for the US Government.

- In your view, how can federal agencies make better use of their performance data to improve management decisions?
- Can you identify any examples of highly-innovative projects to leverage performance data to improve future agency performance?

6. One school of thought holds that agencies may face inherent limitations on their potential to fully implement performance management reforms because their products, services, and program outcomes are difficult to measure quantitatively. Think about diplomacy, for instance.

- To what degree do you believe the mission of federal organizations limits their ability to implement GPRAMA?
- Can you cite examples where this has been shown to be a limiting factor for some agencies?

Appendix O. Location of QCA Among Types of Qualitative Methods



Source: Blatter, Joachim. 2017. "Truth Seeking AND Sense Making: Towards Configurational Designs of Qualitative Methods." *Qualitative and Multi-Method Research*, 15(2): 3. https://doi.org/10.2139/ssrn.3450025. Original figure used with permission of source author and journal; email attached. Arrow added to indicate location of QCA in qualitative methods spectrum.

VIRGINIA TECH

Edgar Hollandsworth <hoya86@vt.edu>

Re: [EXT]Question on 2017 Blatter Article in QMMR

Sebastian Karcher <skarcher@syr.edu> Wed, Jun 30, 2021 at 2:47 PM
To: Blatter Joachim <joachim.blatter@unilu.ch>
Cc: Jennifer Cyr <jmcyr@email.arizona.edu>, Edgar Hollandsworth <hoya86@vt.edu>

Hi everyone,

Joachim retains all copyright to the article, so his permission is all Ed needs to reuse the figure. (It would also be allowed without explicit permission under QMMR's CC license for articles as long as the dissertation is not going to be published by a for-profit publisher.)
Good luck with your dissertation and all the best, Sebastian

Sent from my phone

On Jun 30, 2021 13:42, Blatter Joachim <joachim.blatter@unilu.ch> wrote:

> Dear Ed,
> I am glad to hear that my text and figure is helpful for your research!
> As far as I am concerned, I have certainly nothing against that you are using Figure 1 in your thesis.
> Best
> Joachim
>
> Am 30.06.2021 um 19:23 schrieb Jennifer Cyr <jmcyr@email.arizona.edu>:
>
> Thanks very much, Ed, for your email. Yours is a very good question, and one that I am unsure how to answer. However, I am copying Sebastian Karcher here, who may understand how this kind of permission is granted.
>
> Many thanks, Sebastian, for any help you can provide.
>
> Warmly,
> Jennifer
>
> On Wed, Jun 30, 2021 at 2:19 PM Edgar Hollandsworth <hoya86@vt.edu> wrote:
>
> > External Email
> >
> > Dear Dr Cyr,
> >
> > I am a Public Administration PhD candidate at Virginia Tech working on my dissertation. I enjoyed the article "Truth Seeking and Sense-Making: Towards Configurational Designs of Qualitative Methods" by Joachim Blatter (2017). Figure 1 in that article expresses an important aspect of my study's methodology, and I would very much like to use it as an appendix in my dissertation. Of course, prior approval is required to use copyrighted material, assuming it is. So I would like to request approval to use this. I have cc'd Dr Blatter on this email, in case his approval is also required.
> >
> > Thank you for considering this request.
> >
> > Sincerely,
> > Ed Hollandsworth
> > Ashburn, VA
>
> --
> Jennifer Cyr, PhD,
> (she/hers/ella),
> Associate Professor of Political Science
> Universidad Torcuato di Tella
> Buenos Aires, Argentina
> http://www.jennifercyr.org/
> Red de Politólogas--#NoSinMujeres
> #womenalsoknowstuff

Appendix P. Calibration Thresholds for Outcome and Condition Sets for Fuzzy-Set QCA

Outcome Sets	Scale	Range	Mean	STD	Non-Member	Cross-over	Full Member	Basis for Thresholds
PM (High)	0-7	3-6	4.62	1.2	1.5	3.5	5.5	Crossover: Use of 4 measures or less usually excludes outcome (or alternative) measures; cited as key shortfall in measures adoption (GAO-17-776SP). Full member: Agency uses all legislated GPRAMA measure categories (6) regardless of alternative measures. Non-member: Crossover point for low agency adoption range.
PM (Low)	0-7	0-3	1.25	1.3	0.5	1.5	2.5	Crossover: Point of maximum ambiguity (Ragin 2008). Full member: >2.5 obvious cluster. Non-member: <0.5 obvious cluster and outcome measure criterion.
PI (High)	0-9	4.4-8	5.87	1.2	4.45	5.5	7	Crossover: Point of maximum ambiguity Full member: Clustering of full cases and break below. Non-member: clear clustering and break, and anchor single case as non-member (Freitag and Schlicht 2009).
PI (Low)	0-9	0-2.5	1.5	0.9	0.5	1.5	2.25	Crossover: Point of maximum ambiguity. Full member: Clustering of cases above large middle group. Non-member: Cluster of two agencies at zero with clear break from middle cluster.

Condition Sets	Scale	Range	Mean	STD	Non-Member	Cross-over	Full Member	Basis for Thresholds
P&R	0-10	2-6	3.65	1.3	2.25	3.25	4.5	Crossover: Point of maximum ambiguity Full and Non-member: Clear clustering and breaks. For Non-member: Anchor single case as non-member.

C	0-7	2-6.5	4.09	1.3	2.5	3.5	4.25
L	0-8	1-7	4.62	1.9	2	4.2	5.5
X	0-6	1-5	3.51	1.3	2	3.8	4.75
M	0-9	5-8	6.56	1.1	5.25	6.5	7.75

P&R Political Influences and Resource Dependency
C Agency Culture of Accountability and Empowerment
L Credible Leadership Commitment
X Capacity for Evaluation of Performance
M Measurement System Maturity

Crossover: Point of maximum ambiguity. Full and Non-member: Clear clustering and breaks. Anchor single case as non-member.

Crossover: Point of maximum ambiguity. Full and Non-member: Clear clustering and breaks. Anchor single case as non-member.

Crossover: Point of maximum ambiguity. Full and Non-member: Clear clustering and breaks. Anchor single case as non0-member.

Crossover: Point of maximum ambiguity. Full and Non-member: Clear clustering and breaks. Anchor single case as non-member.

Appendix Q. Evidence from Interviews with Performance System Managers at HUD

Conditions (Survey Score/Federal Mean)	**Evidence**
Structure and Processes Supporting GPRAMA Reporting (NA)	• PIO/CFO functions merged • COO separate from Deputy Secretary • Deputy PIO leads strategic planning and performance team and coordinates across highest-levels of HQ and program offices • PMS governance decisions taken at "highest levels" • Tiny agency (6,500-7,000 employees) but huge budget executed via grantees at state and local levels; HUD is more like a "bank" • Originally emulated CityStat model; moving towards project management use of performance information vice accountability • Monthly task force meetings (vice quarterly) • Applied project management paradigms during Carson era • Measures in program offices, not well-connected to HUD Strategic Plan/GPRAMA reports; exception: Agency Priority Goals (APGs)
External Politics and Organizational Dependency on Resources (3.00/3.65)	• External partners include industry associations and public housing authorities (PHAs), are very numerous (more partners than program offices) but have "little say" in HUD PMS. • Secretary and Deputy Secretary personally decide how to implement GPRAMA and OMB guidance; take notice of partner interactions with Congress on HUD programs • Rapid turnover of political appointees (<24 months average) leads career officials to "wait them out" on implementing reforms
Accountability and Empowerment in Organizational Culture (2.00/4.09)	• Fear of accountability-Some executives try to keep programs out of strategic plan to avoid scrutiny and potential embarrassment for being called out for low performance in front of peer executives. • Strategic Plan measures integrated into SES performance reviews. • Cultural view of GPRAMA-just a compliance exercise • Aging workforce (average age 56) resists PMS changes but is passionate about mission (PMS seen as a distractor, not a helper) • Workforce suspicious of Congress; sees itself as a recurring target for criticism and attempts to dissolve – led to sense of "resiliency" • Cited entry level workforce retention challenges caused by image of HUD as an agency that resists change due to seniority of workforce.
Credible Commitment of Senior Leaders in the Organization (1.00/4.62)	• Leadership styles and preferences for PM changed between recent Secretaries, reduced use of outcome-based measures (e.g., HUDStat). • Manager views of credible commitment strongly depend on signals (actions and decisions) from Secretary and Deputy Secretary. • Key challenge is getting executives to voluntarily participate in PMS.

Organizational Capacity to Evaluate Performance Data (4.00/3.50)	• Reorganizations in past 4 years cut PMS staff by 90 percent, handicapping HUD's planning function. • Program offices now developing their own strategic plans, learned from prior enterprise level strategic planning successes. • Office of Policy Research has capacity to formulate new performance measures when needed. • External state and local partners have very little or no capacity to evaluate program performance, who execute almost all of HUD's budget; perception that investments to make possible would impact resources need to perform. • "In order to improve anything, we need higher levels of support."
Maturity of the Organization's Performance Measurement System (6.00/6.55)	• Performance measure standardization exists, but only within program offices, not at enterprise level. • HQ-level strategic planning and performance team "picks and plays" measures from program offices for GPRAMA reports. • Performance info used to help justify budgets related to APGs. • Lack of IT interoperability between program offices cited as one cause for lack of standardization of measures at enterprise level • Cited the need for improved integration between strategic plan, budget, and "everything else;" APGs help bring all together. • Innovation: HUD links enterprise fraud and risk management system with strategic planning process – risks now considered in plans.
Noteworthy Agency-Specific Characteristics Affecting GPRAMA Implementation	• Confirmed results of fixed effects quantitative model that PI use for setting performance goals in HUD was well below the federal agency average. • Noted PI use in HUD Headquarters has not migrated down to the program office level yet. Projected institutionalization at program level in 5-15 years.

Appendix R. Evidence from Interviews with Performance System Managers at NASA

Conditions (Survey Score/Federal Mean)	**Evidence**
Structure and Processes Supporting GPRAMA Reporting (NA)	• PIO is director of Strategic Investments Division (SID); provides enterprise-level guidance and ensures compliance and reports to CFO. • All goals and strategic objectives are owned by mission directorates and not changed by SID; SID provides guidance and advice only. • Mission directorates develop their GPRAMA measures internally under administration of a performance officer. • Development of measures is a collaboration between mission directorates and the A-Suite. • NASA has a robust system of program performance reviews and measures separate from GPRAMA, but it feeds into GPRAMA reporting. • Program data/performance reviewed from project leads up to A-Suite. • Baseline Performance Reviews (BPR) conducted monthly with all executive leaders, includes GPRAMA measures and other topics • NASA sees itself as an R&D agency and so measures its performance with non-GPRAMA frameworks appropriate for R&D management. • Standardization of program performance measures is commonality of measurement types between mission offices, not at enterprise level. Cited real challenges standardizing internal mission support measures.
External Politics and Organizational Dependency on Resources (2.5/3.65)	• Congressional influence and feedback on GPRAMA are nonexistent. • OMB is only outside stakeholder with interest in GPRAMA in NASA; PSMs engage with OMB one-on-one on strategic plan, goals, etc. • OMB influence not political, but historically sought to exclude mission support performance (e.g., IT, space, comms, etc) from GPRAMA strategic plans (not mission-related); views now changing. • NASA's R&D identity drives what stakeholders care about. • Much of NASA's budget is discretionary – budget and strategic plan are constantly changing, and change makes SID work challenging. • Political influence can skew NASA planning: Trump Administration directed nothing climate-related was to be in NASA strategic plan; as result, NASA had funded "Zombie" programs that were not evaluated. • Legislative policy preferences can change strategic plan; example – manned space flight program interest re-emerging.
Accountability and Empowerment in Organizational Culture (6.50/4.09)	• Argued GPRA was "check the block" exercise but GPRAMA is bringing cultural changes to NASA. • NASA executives well aware of their GPRAMA responsibilities and are highly responsive to BPR meetings and put program data in front of NASA senior leadership.

	• SID believes line managers have made performance reviews part of routine practices down to project level – but most PI is used for internal tracking vice external accountability (e.g., milestones, cost, schedule, etc). Strong *program-level* accountability. • SES performance appraisals include strategic plan performance goals; GS appraisals do also indirectly (based on their projects).
Credible Commitment of Senior Leaders in the Organization (6.00/4.62)	• "Senior leadership engagement has really helped NASA to implement GPRAMA." • PIO and DPIO are engaged with the A-Suite, look for opportunities to chart ways ahead. • Senior leaders very engaged in strategic planning – COO personally leads kickoff meetings. • Program offices responded positively to perceptions senior leaders were committed to performance measurement; are sensitive to COO reviews of their performance reports.
Organizational Capacity to Evaluate Performance Data (3.50/3.50)	• Senior leaders want high profile measures but lack skill to write good performance measures for APGs. • SID lacks resources to conduct analysis of performance; what analysis takes place is at program office level. • Biggest problem is budget constraints. • Discarded previous clunky software but do not have resources to replace it; hobbling along using Excel. • Noted staffing constraints that create difficulties. No funds to hire more people. • Some mitigation from other branches in SID which analyze NASA with a different view of performance meaning.
Maturity of the Organization's Performance Measurement System (5.50/6.55)	• Several years ago, NASA had challenges operationalizing outcome measures; engagement with OMB and deep analysis fixed that. • Until 2012, no NASA measures included success criteria; "It took years to perfect that." • History of "loose" performance measurement practices but NASA has tightened up standard and now measures are more reliable and valid. • Improved measure quantification to enable trending of data. • Annual targets for measures were added for the first time in 2020. • Performance reports include both stoplight charts and trend/target data to address needs of senior executives and program managers. • Performance measures improve management effectiveness, but there is no evidence they reduce costs.
Noteworthy Agency-Specific Characteristics Affecting GPRAMA Implementation	• Quantitative analysis of 2017 GAO manager survey data indicated NASA was far above federal average for adopting outcome measures and for use of PI for problem identification, corrective actions, contract management, and program consolidation, but below average for use of PI for program strategy development. • Prior to 2017, NASA learned to create outcome variables and worked closely with OMB to implement A-11 requirements. Regularly

	refreshing strategic plans increases the focus on outcome goals and measures (set 2–5-year windows for performance goals and indicators). • PI use responses on GAO survey likely result of NASA line managers using project and program level performance data, not so much GPRAMA measures. Example: program decisions driven by cost and schedule milestones more than GPRAMA metrics. • Use of PI for program strategy got a boost after 2017 as NASA has learned to incorporate success criteria into strategic objectives.

Appendix S. Evidence from Interviews with Performance System Managers at Treasury

Conditions (Survey Score/Federal Mean)	**Evidence**
Structure and Processes Supporting GPRAMA Reporting (NA)	• COO is Deputy Secretary. PIO is Asst Secretary for Management. DPIO is also Director of Office of Strategic Planning and Performance Improvement (OSPPI). • Leaders appointed for agency goals and strategic objectives. • 30-40 performance contacts in bureaus and offices – performance community of practice. • Three synchronized layers: strategic planning: 5-year cycle for outcomes and strategic objectives, annual budget formulation, and ongoing strategic management routines. • Well-planned quarterly cycle of meetings: previous FY review (fall), annual strategic review (winter), focused issues review (spring), future years budget preparation (summer) • Performance contacts in bureaus reporting greater line manager use of PI than 8 years ago; managers use PI but may not be aware of a connection to GPRAMA performance reporting; they use metrics for internal use, not reported to the public. • Cited good standardization in the process for developing and validating measures; less uniformity in departmental processes for reporting, analyzing or using PI (depends on approach of analyst). • Published a policy on measuring and reporting organizational performance. • Conducts an annual performance measures assessment – value-added routine to validate relevance and quality of measures.
External Politics and Organizational Dependency on Resources (3.00/3.65)	• Very little external influence on GPRAMA implementation except for OMB, which provides "a lot of latitude." • No evidence of external political or resource dependence on Treasury institutionalization of GPRAMA. • Deputy Secretary, as COO, sees GPRAMA as a "management tool" and thus sees himself as a key stakeholder; his preferences shape institutionalization decisions. • PSMs have observed some managers being unfairly held to account for sub-par performance for reasons beyond their control; rare but has had a "chilling effect" when it happened • Community Development Financial Institutions Fund has been a "political football" in terms of annual budget uncertainty
Accountability and Empowerment in Organizational	• Managers collaborate and are generally held accountable for achieving results. • History of success with performance management; e.g., classic case of paperless payment processing, went from 40-90 percent paperless in 4

Culture (3.00/4.09)	years during early years of GPRAMA; current leaders working with IRS to reduce total number of checks to beneficiaries via e-deposit encouraged by COVID legislation.
Credible Commitment of Senior Leaders in the Organization (5.00/4.62)	• Overall very strong credibility of senior leader commitment (5/7 survey responses "agree" or "strongly agree" • Leaders continuously communicate expectations for continuous improvement, more effective practices, and goal-setting. • Treasury maintains an index of performance culture, tracks the trend over time. • Very high level of senior executive participation in QPRs; about 140. • PSMs observed very high quality of performance discussions in QPRs. • Annual Federal Employee Viewpoint Survey also points to strong senior leader commitment perceptions by Treasury's workforce.
Organizational Capacity to Evaluate Performance Data (1.00/3.50)	• Performance staff is "thin on the ground." • Measurement resources are sufficient for legal requirements, but not the intent to deeply institutionalize performance measurement. • To expand capacity further would require leaders and program managers to obtain "a higher level of literacy around performance measure development." • Treasury has programs to educate managers on performance measurement and is making progress.
Maturity of the Organization's Performance Measurement System (5.00/6.55)	• PSMs could not verify that employees share the same definition of performance management concepts. • Measurement system is linked by business practices to individual performance evaluations and internal management improvement consulting and performance coaching teams.
Noteworthy Agency-Specific Characteristics Affecting GPRAMA Implementation	• Quantitative analysis of 207 GAO survey of federal managers suggested Treasury managers were challenged in using PI for setting program priorities and using PI for managing contracts. • Treasury reviewed 2017 GAO federal manager survey and sent reclama to GAO with concerns over survey assumptions and methods. • Primary concern was 2017 survey results better reflected IRS than Treasury as a whole because 90 percent of Treasury managers work in IRS[91]; results were not representative of Treasury as whole because IRS had implementation challenges in 2017 and there is still a lingering stigma in the IRS against performance management.

[91] According to the *Treasury Strategic Plan* for FY2018-2022, the IRS is one of seven bureaus that report to Treasury (https://home.treasury.gov/system/files/266/Treasury-Strategic-Plan-FY-2018-2022.pdf).

Appendix T. Evidence from Interviews with Performance System Managers at DHS

Conditions (Survey Score/Federal Mean)	**Evidence**
Structure and Processes Supporting GPRAMA Reporting (NA)	• COO is the Deputy Secretary; PIO is the Chief of Office of Program Analysis and Evaluation (PA&E); PA&E has 5 performance analysts, aligned under CFO. • PA&E sets performance management policies/procedures, orchestrates quarterly performance reviews (QPRs), APGs, and strategic reviews. • DHS performance community includes about 40 performance officers in DHS elements plus PIOs in components; meets quarterly. PIO council meets monthly. Component officers work policy and other matters, not just performance measurement. • PA&E not allowed to input to Strategic Plan; as a result, there are "too many overlapping objectives" and it is hard to align GPRAMA performance measures (235) with the Strategic Plan. • GPRAMA is not socialized deeply at lower levels, so DHS has two distinct types of performance measurement: "above the waterline" (public-facing) and "below the waterline" (internal use). • At program level, below the "water line," leaders generally use after-action reports and mission analysis, not performance measures. • DHS PSMs see it as role of Chief Human Capital Officer to create incentives for employees to implement GPRAMA.
External Politics and Organizational Dependency on Resources (2.80/3.65)	• PA&E relies on OMB direction (especially in budgetary pass-back guidance) to get DHS components to measure their performance. • Political leaders control public messaging; have blocked publication of some poor performance results. • Legal requirement of GPRAMA is the only reason DHS senior leaders budget resources for performance management; they would not do it without the law. • PSMs did not see involvement by political leaders outside of DHS. • Focus on DHS budget and resources commands most of DHS leaders' attention: "money is sexy, performance is boring" to them.
Accountability and Empowerment in Organizational Culture (4.60/4.09)	• PSMs perceive DHS Director does not really use his authority to make components "get serious" about GPRAMA and allows components to recuse themselves from process without justification. • PA&E prepares scorecard for CFO every four months to evaluate components; enables CFO to hold components accountable but little evidence CFO does so. • Accountability at executive level emphasized by including strategic plan outcome measures in SES individual performance plans, incentivized with bonuses. • Senior executives work with PA&E to prepare for strategic reviews.

	• Sone components feel they lack the authority to hold sub-components accountable for weak performance against APGs. • "Performance is not high on the list of priorities for most managers." • PSMs felt managers need to gain more experience using performance measurement. • Managers associated with successes recognized at strategic reviews. • Culture of blame avoidance for not achieving benchmarks by avoiding attribution of outcomes to programs and discounting benchmarks based on incomparability between diverse DHS mission areas.
Credible Commitment of Senior Leaders in the Organization (4.40/4.62)	• Rapid turnover in Secretaries and political appointees ("a lot of leadership churn") made it hard to get appointees interested in PI use. • DHS Secretary and Deputy appear to be preoccupied with press reports about DHS and effects on agency's image, so not focused on improving PMS. • Some unspecified "personality" and "power and control" issues create barriers between DHS HQ and components – hinders institutionalization of GPRAMA. • DHS meetings on agency priority goals structured to meet "letter of the law" but not see as useful internal management tool. • Customs and Border Patrol leaders are more credible; meet "routinely" with their managers to discuss performance. • Managers have concerns about using PI if they believe it will be used "punitively" against them.
Organizational Capacity to Evaluate Performance Data (4.40/3.50)	• Sufficient capabilities to "stay compliant" with law. • DHS runs an online tool for components to submit quarterly data for QPRs. • Due to lack of resources, DHS does not do enough process improvement work like BPR, Six Sigma, root cause analysis, etc. • PA&E is "bogged down just doing GPRAMA requirements." • "It takes six months to do a strategic review" • Pockets of program process improvements exist in some components. • Use a "standard database" to track data and trends. • DHS invested in expansion of Internal Controls Program in 2009 to develop tools and documentation and measure V&V – did improve data quality with good results for budget outcomes. • "Big-E" evaluation (non-program) capabilities just now "standing up." • "Maybe we need a performance capacity assessment capability."
Maturity of the Organization's Performance Measurement System (7.00/6.55)	• GPRAMA measure definitions are standardized using a form; form also identifies sources of information to be used and who collects data – makes measures more "reliable." • But no standard measures in DHS because each component's mission is different; inspection criteria vary between components. • Use benchmarks and historical data to compare trend data; some benchmarks based on customer service/satisfaction indices.

	• DHS leverages Internal Controls Program for *ex post* measures validation by expert review team with PhDs and contractors. • Below "water line," experts support non-GPRAMA measures with a handbook, clear criteria, and audit-like checks to improve data quality. • PA&E runs a V&V review semi-annually to validate measures and identify issue with methodology with help from outside contractor.
Noteworthy Agency-Specific Characteristics Affecting GPRAMA Implementation	• PSMs noted DHS was still a relatively new agency and was still maturing. Felt managers needed to gain more experience with performance management. • DHS has an internal performance community of practice. It, too, is developing its knowledge through training meetings every 2-3 months. • PA&E holds performance measurement training for components but manager interest is tepid as turnout is "not always strong."

Appendix U. Evidence from Interviews with Performance System Managers at SBA

Conditions (Survey Score/Federal Mean)	**Evidence**
Structure and Processes Supporting GPRAMA Reporting (NA)	• Small staff of about 3,000 full-time equivalent positions. • Deputy Administrator serves as the COO. The PIO is the CFO. • Deputy PIO is Director of Office of Program Performance Analysis and Evaluation (OPPAE) and reports to CFO. Leads two divisions: one for GPRAMA and one for Evidence Act (program evaluation). • Noted strategic goals and objectives are integrated into employee performance plans. • Confirmed SBA uses a wide range of performance measures. • Agency-level Quarterly Performance Reviews with program managers. • PM processes are standardized by OPPAE working with program offices annually on QPRs, strategic reviews, review of measures, strategy updates, performance plans, and measuring progress. • Performance management closely integrated with other decision tools like program evaluation, regulatory analysis, performance reports.
External Politics and Organizational Dependency on Resources (2.00/3.65)	• Congressional authorization and appropriations committee have oversight of SBA's budget, but allocate using incremental methods, not based on performance measures. • SBA's political leadership team is "very engaged" with performance management system in a constructive way.
Accountability and Empowerment in Organizational Culture (4.00/4.09)	• Small staff at SBA supports an entrepreneurial mindset and informal relationships across the organization. • Small staff also means SBA must be "nimbler in how we approach our processes;" so they seek cost-effective measures. • SBA senior leaders hold program managers accountable at QPRs. • Deputy Administrator, as COO, chairs QPRs; PSMs perceive this ensures accountability of program offices.
Credible Commitment of Senior Leaders in the Organization (7.00/4.62)	• PSMs perceive "a high level of commitment" by SBA senior leaders. • OPPAE adds credibility to senior commitment by communicating expectations for constructive collaboration on performance improvement, not just tasking to collect performance data. • Informal and trusting relationship between OPPAE and program managers is an essential condition for successful GPRAMA institutionalization in SBA.
Organizational Capacity to Evaluate	• PMS staff are adequate, experienced, and highly knowledgeable about procedures for performance management and evaluation. • PMS staff are also knowledgeable about program evaluation.

Performance Data (4.00/3.50)	• SBA developed and acquired useful tools such as digital dashboards to analyze program-level data. • Innovation in PMS noted in early adoption of program evaluation capability even before 2018 Evidence Act, creation of dashboards. • PSMs believe SBA capacity "goes beyond GPRAMA requirements." • PSMs note program offices willingly provide requested performance data to OPPAE.
Maturity of the Organization's Performance Measurement System (8.00/6.55)	• Overall PSM team assessment: The SBA PMS is "good and mature." • New measurement practices for strategic reviews build on already successful legacy practices. • Innovated using templates for program offices to use to report performance and inform SBA senior leaders. • Incorporates risk management analysis into strategic reviews.
Noteworthy Agency-Specific Characteristics Affecting GPRAMA Implementation	• PSMs attribute exceptionally high agency-specific adoption of output measures to nature of SBA programs (loan processing and government contracting certification) measured based on their outputs. • Low use of PI for corrective actions reflects line manager perceptions that employee performance cannot be fairly evaluated because SBA programs are non-mandatory: the economy and decisions made by private lenders and companies determine program outcomes. • Low use of PI for Contract Management reflects legal restrictions on small business set aside goals and sub-goals, which limits use of PI for contracting decisions (73 percent of SBA contracts go to small businesses).