



4-2004

Implementing Performance Measurement Innovations: Evidence From Government

Ken S. Cavalluzzo

Christopher D. Ittner
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/accounting_papers

 Part of the [Accounting Commons](#)

Recommended Citation

Cavalluzzo, K. S., & Ittner, C. D. (2004). Implementing Performance Measurement Innovations: Evidence From Government. *Accounting, Organizations and Society*, 29 (3-4), 243-267. [http://dx.doi.org/10.1016/S0361-3682\(03\)00013-8](http://dx.doi.org/10.1016/S0361-3682(03)00013-8)

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/accounting_papers/40
For more information, please contact repository@pobox.upenn.edu.

Implementing Performance Measurement Innovations: Evidence From Government

Abstract

Using data from a government-wide survey administered by the US General Accounting Office, we examine some of the factors influencing the development, use, and perceived benefits of results-oriented performance measures in government activities. We find that organizational factors such as top management commitment to the use of performance information, decision-making authority, and training in performance measurement techniques have a significant positive influence on measurement system development and use. We also find that technical issues, such as information system problems and difficulties selecting and interpreting appropriate performance metrics in hard-to-measure activities, play an important role in system implementation and use. The extent of performance measurement and accountability are positively associated with greater use of performance information for various purposes. However, we find relatively little evidence that the perceived benefits from recent *mandated* performance measurement initiatives in the US government increase with greater measurement and accountability. Finally, we provide exploratory evidence that some of the technical and organizational factors interact to influence measurement system implementation and outcomes, often in a complex manner.

Disciplines

Accounting

**IMPLEMENTING PERFORMANCE MEASUREMENT INNOVATIONS:
EVIDENCE FROM GOVERNMENT**

Ken S. Cavalluzzo
Georgetown University

Christopher D. Ittner
University of Pennsylvania

January 2003

We are grateful to the United States General Accounting Office for providing access to the data used in this study. We appreciate the assistance of Tom Beall in obtaining the data and thank Jennifer Ciralsky for excellent research assistance. We also thank Shannon Anderson, David Cooper, Ranjani Krishnan, Bill Lanen, Cathy Tinsley, Jerry Zimmerman, two anonymous reviewers, and participants at the University of Rochester, Washington University, and the 2002 AAA Management Accounting Conference for valuable comments and discussion. The conclusions in this paper are those of the authors and not those of the General Accounting Office.

IMPLEMENTING PERFORMANCE MEASUREMENT INNOVATIONS: EVIDENCE FROM GOVERNMENT

Abstract

Using data from a government-wide survey administered by the U.S. General Accounting Office, we examine some of the factors influencing the development, use, and perceived benefits of results-oriented performance measures in government activities. We find that organizational factors such as top management commitment to the use of performance information, decision-making authority, and training in performance measurement techniques have a significant positive influence on measurement system development and use. We also find that technical issues, such as information system problems and difficulties selecting and interpreting appropriate performance metrics in hard-to-measure activities, play an important role in system implementation and use. The extent of performance measurement and accountability are positively associated with greater use of performance information for various purposes. However, we find relatively little evidence that the perceived benefits from recent *mandated* performance measurement initiatives in the U.S. government increase with greater measurement and accountability. Finally, we provide exploratory evidence that some of the technical and organizational factors can interact to influence measurement system implementation and outcomes, often in a complex manner.

IMPLEMENTING PERFORMANCE MEASUREMENT INNOVATIONS: EVIDENCE FROM GOVERNMENT

1. Introduction

Performance measurement issues are receiving increasing attention as organizations attempt to implement new measurement systems that better support organizational objectives. While many of these initiatives are in the private sector, recent efforts to improve governmental performance have also placed considerable emphasis on performance measurement as a means to increase accountability and improve decision-making (Ittner and Larcker, 1998). Indeed, Atkinson, Waterhouse, and Wells (1997) note that government agencies are at the forefront of efforts to implement new, more strategic performance measurement systems. The Government Performance and Results Act of 1993, for example, requires United States executive branch agencies to clarify their strategic objectives and develop results-oriented measures of progress towards these objectives. Similar initiatives have been launched in Australia, Canada, New Zealand, the United Kingdom, and other countries (Smith, 1993; Hood, 1995; Atkinson and McCrindell, 1997).

This study draws upon the information systems change, management accounting innovation, and public sector reform literatures to examine some of the factors influencing the implementation, use, and perceived benefits of results-oriented performance measurement systems in the U.S. government. Small-sample studies in both the public and private sectors identify a number of potential impediments to the successful implementation of performance measurement innovations (e.g., GAO, 1997a; Gates, 1999). These impediments include identifying appropriate goals in environments characterized by multiple and conflicting objectives, measuring performance on hard-to-

evaluate or subjective goals, overcoming deficiencies in information systems, providing incentives for employees to use the information to improve performance, and achieving management commitment to the new systems. Because many of these problems are present across the public and private sectors, the broad-scale implementation of new performance measures in the U.S. government provides an attractive setting to examine some of the factors influencing the success or failure of measurement system innovations.

Consistent with information system and management accounting change models (e.g., Kwon and Zmud, 1987; Shields and Young, 1989), we find that organizational factors such as top management commitment to the use of performance information, the extent of decision-making authority delegated to users of performance information, and training in performance measurement techniques have significant positive influences on measurement system development and use. However, we also find that technical issues play an important role in performance measurement system implementation and use. In particular, difficulties selecting and interpreting appropriate performance metrics in hard-to-measure activities are a major impediment to measurement system innovation. Data limitations, such as the inability of existing information systems to provide necessary data in a valid, reliable, timely, and cost effective manner, also deter the use of performance information for accountability and performance evaluation. Technical issues such as these appear to play a much more important role in the implementation of performance measurement systems than they do in cost system implementation (e.g., Shields, 1995; Krumwiede, 1998, Anderson and Young, 1999).

The extent of performance measurement and accountability are positively associated with the use of performance information for various purposes, consistent with

claims that improved performance information and incentives for achieving results can support governmental decision-making. However, we find relatively little evidence that the perceived benefits from recent *mandated* performance measurement initiatives in the U.S. government increase with greater measurement and accountability. The latter results support institutional theories that claim systems implemented to satisfy external requirements are less likely to influence internal behavior than are those implemented to satisfy the organization's own needs.

The remainder of the paper contains five sections. Section 2 provides an overview of recent performance measurement initiatives in the U.S. government and develops our hypotheses. Section 3 discusses our sample, followed by descriptive statistics on the variables used in our study in Section 4. Results are presented in Section 5. Section 6 offers our conclusions.

2. Background and Hypotheses

2.1 Performance Measurement Initiatives in the U.S. Government

During the 1990s, the U.S. government began enacting several major initiatives to promote a performance-based approach to the management and accountability of federal activities, including the Chief Financial Officers Act, the National Performance Review, and the Government Performance and Results Act. The stated goals of these initiatives are twofold: (1) to increase Congressional oversight and foster greater accountability for achieving results, and (2) to enhance "performance-based" decision-making by implementing information systems that supplement traditional input-oriented performance measures (e.g., expenditures and staffing levels) with measures focused on

results (e.g., output quantity, quality, and timeliness) and the achievement of strategic objectives.

The most important initiative is the Government Performance and Results Act of 1993 (hereafter, GPRA). The GPRA requires managers of each government activity (i.e., project, program, or operation) to clarify their missions and strategic objectives and to measure relevant outputs, service levels, and outcomes for each activity in order to evaluate performance toward these objectives (U.S. Senate, 1992; GAO, 1997b). Pilot GPRA implementations began in fiscal 1994, with all major agencies required to submit performance goals and indicators for each of their individual activities by fiscal 1997.

The GPRA and related initiatives in other countries are based on the assumption that mandated reporting of results-oriented, strategic performance indicators can improve governmental efficiency and effectiveness by increasing the accountability of public managers (Osborne and Gaebler, 1993; Atkinson and McCrindell, 1997; Jones and McCaffery, 1997). According to the Governmental Accounting Standards Board's Concept Statement No. 2, public sector accountability represents the duty for public managers to answer for the execution of their assigned responsibilities, and for citizens and their elected or appointed representatives to assess performance and take actions by allocating resources, providing recognition or rewards, or imposing sanctions based on the managers' results. By making public officials, legislative bodies, and the public more informed about the behavior of government managers and the results of their actions, the performance measurement initiatives are intended to improve the allocation of

government resources and promote governmental efficiency and effectiveness through improved performance-based decision-making (Flynn, 1986; Scott, 1987).¹

2.2 Determinants of Measurement System Implementation and Success

Prior studies on information system change, management accounting innovation, and public sector reform have identified a number of factors that are expected to influence the implementation and success of performance measurement initiatives such as the GPRA. These factors include *technical* issues, such as the ability of existing information systems to provide required data and the extent to which organizations can define and develop appropriate measures, and *organizational* issues, including management commitment, decision-making authority, training, and legislative mandates (e.g., Kwon and Zmud, 1987; Shields and Young, 1989).

Drawing upon this literature, we employ the conceptual model in Figure 1 to investigate the relations among these factors, the extent of measurement system development, and the stated objectives of governmental performance measurement initiatives (i.e., greater accountability for achieving results, enhanced decision-making, and, ultimately, improved government efficiency and effectiveness). The following sections develop our hypotheses regarding the expected relations between the various technical and organizational factors and the extent of measurement system implementation and outcomes.

¹Many observers argue that the government performance measurement initiatives are emulating the private sector by adopting similar mechanisms for controlling principal-agent problems (Smith, 1990, 1993; Mayston, 1993). See Rose-Ackerman (1986), Tirole (1994), and Dixit (1997) for theoretical studies focused on the applicability of principal-agent models of management control practices in the public sector.

2.2.2 Information System Capabilities

Kwon and Zmud's (1987) review of the information technology (IT) implementation literature indicates that some of the key factors influencing implementation success are technological issues. These issues include the compatibility of the new system with existing systems, system complexity, and the system's relative improvement over existing systems (e.g., accuracy and timeliness). Accounting researchers have drawn upon this literature to argue that the success of management accounting innovations should also be a function of the current information system's capabilities. Krumwiede (1998), for example, suggests that organizations with higher quality information systems may be able to implement new measurement systems more easily than organizations with less sophisticated information systems because measurement costs are lower, leading to a positive relation between current information system capabilities and implementation success. Conversely, managers who are generally satisfied with the information from the existing system may be reluctant to invest the necessary resources in the new system, leading to a negative relation.

Academic studies provide mixed evidence on the influence of information system issues on accounting system innovations. Shields (1995) finds *no* association between successful implementation of activity-based costing (ABC) and technology (i.e., type of software or stand-alone vs. integrated system). Anderson and Young (1999) find that the perceived quality of the existing information system is *negatively* related to management's evaluation of ABC success. Krumwiede (1998) reports a *positive* association between the strength of the existing information system and an organization's decision to undertake more advanced stages of ABC adoption, but not with earlier stages.

Surveys of performance measurement innovations in the private sector, on the other hand, indicate that information system problems represent a major impediment to implementation success. Many of these problems relate to the ability of existing information systems to provide required data in a reliable, timely, and cost effective manner. Gates' (1999) study of strategic performance measurement (SPM) systems concludes that most companies' information technologies (IT) are limited in their ability to deliver rapid and consolidated results for analysis. In addition, nearly 60% of his respondents avoid using certain strategic performance measures due to limitations in their IT systems, 22% do not believe their IT systems capture data sufficiently, and 57% are forced to capture at least some SPM information manually. A survey of balanced scorecard users by Towers Perrin also finds that the lack of highly-developed information systems is a problem or major problem in 44% of scorecard implementations (Ittner and Larcker, 1998).

Small-sample field studies in the public sector report similar results (Jones, 1993; GAO, 1997a). These studies suggest that information system problems in government organizations are compounded by the need to use data collected by other organizations (e.g., other federal organizations, state and local agencies, and non-government recipients of federal funds) and difficulties ascertaining the accuracy and quality of this data. Kravchuk and Schank (1996) conclude that the intergovernmental structure of many programs results in serious measurement problems when the information systems used by different organizations vary in terms of data definitions, technology, ease of accessibility, and amount of data retained. If these information system limitations prevent managers from receiving timely and reliable data, the performance measurement system's use for

accountability and decision-making purposes is likely to be limited (Jones, 1993; Kravchuk and Shank, 1996).

These issues prompt our first hypothesis:

H1: Performance measurement development and outcomes are negatively associated with problems obtaining necessary data in a reliable, timely, and cost effective manner.

2.2.3 Selecting and Interpreting Performance Metrics

A second technical issue highlighted in the performance measurement literature is the ability to define and assess metrics that capture desired actions and outcomes.² In many public and private sector settings, employees carry out many tasks that are difficult to accurately evaluate using objective, quantifiable performance metrics (e.g., basic research and development activities). In these settings, theoretical studies indicate that the implementation and effectiveness of performance measurement systems are likely to be low (e.g., Holmstrom and Milgrom, 1991), with greater emphasis placed on subjective, qualitative judgments when evaluating performance than on quantitative performance metrics (e.g., Prendergast, 1999).

Surveys of private sector measurement practices indicate that problems identifying and measuring appropriate performance metrics represent significant impediments to system success. Gates (1999) finds that the leading roadblocks to implementing strategic performance measurement systems are avoiding the measurement of “difficult-to-measure” activities (55% of respondents), measuring “the right things

² The terms performance metric and performance measure are interchangeable. We refer to performance metrics when discussing the identification, development, and interpretation of specific performance measures for evaluating managerial performance or aiding decision-making. We refer to performance measure development or performance measurement systems more generally as a collection of performance metrics that are reported on a regular basis through the organization’s information systems.

wrong” (29%), and measuring “the wrong things right” (29%). Similarly, the Towers Perrin survey of balanced scorecard users finds that 45% of respondents view the need to quantify qualitative results to be a major implementation problem (Ittner and Lareker, 1998).

In the public sector, empirical and theoretical studies indicate that problems selecting appropriate metrics and interpreting results often stem from four features common to many federal programs (as well as many activities in the private sector): (1) the complicated interplay of federal, state, and local government activities and objectives, (2) the aim to influence complex systems or phenomena whose outcomes are largely outside government control (e.g., programs that attempt to intervene in ecosystems, year-to-year weather, or the global economy), (3) missions that make it hard to develop measurable outcomes (e.g., prevention of a rare event such as a presidential assassination), to attribute results to a particular function (e.g., reductions in unemployment), or to observe results in a given year (e.g., basic scientific research), and (4) difficulties measuring many dimensions of social welfare or other governmental goals (e.g., Tirole, 1994; Dixit, 1997; GAO, 1997a). The GAO (1997a) argues that problems such as these can force organizations to develop performance metrics that are incomplete or uninformative in order to meet the GPRA’s reporting requirements, with limited use of the resulting metrics for decision-making and accountability purposes.

These issues lead to our second hypothesis:

H2: Performance measurement development and outcomes are negatively associated with difficulties selecting and interpreting appropriate performance metrics.

2.2.4 Management Commitment

While technical factors are expected to significantly influence the implementation of performance measurement innovations, their impact may be secondary to that of organizational factors (Shields and Young, 1989). Shields (1995), for example, argues that top management support for the innovation is crucial to implementation success because these managers can focus resources, goals, and strategies on initiatives they deem worthwhile, deny resources to innovations they do not support, and provide the political help needed to motivate or push aside individuals or coalitions who resist the innovation.

The information system change literature also highlights the role of top management support in creating a suitable environment for change, influencing users' personal stakes in the system, and increasing the appreciation of others for the potential contribution of the system to meeting organizational objectives (e.g., Manley 1975; Schultz and Ginzberg, 1984; Doll, 1985). Consequently, employees who perceive strong support for the system by top management are more likely to view the change favorably (McGowan and Klammer, 1997).³ Top management commitment is therefore expected to influence both the extent to which employees feel accountable for results and their use of the information for decision-making.

The need for strong top management commitment to performance measurement is recognized in the government reform literature. The GAO (1997b) argues that results-oriented performance measurement initiatives will not succeed without the strong

³ A positive relation between top management's commitment to using new performance measures and their use by lower-level managers can also be explained by contagion effects, which represent the spread of a particular process or paradigm from one level of management hierarchy to the next (Macintosh, 1985).

commitment of the U.S. federal government's political and senior career leadership. However, Flynn (1986) notes that performance measurement initiatives are part of government efforts to cut expenditures. The implication is that efficiency improvements will lead to lower budgets, reducing incentives for top management to support performance measurement efforts. Jones (1993) adds that U.S. executive branch officials do not want to aid Congressional oversight committees in the micro-management of executive agencies, or to assist Congress in gaining leverage over the president and his cabinet appointees. Consequently, there may be little reason for top agency management to support performance measurement efforts. Jones and McCaffery (1997) also find that Congressional knowledge of and interest in performance measurement initiatives are low, and argue that Congress, which is motivated by short-term re-election concerns, is institutionally incapable of making long-range decisions based on the performance measures mandated by the GPRA. As a result, legislators' commitment to the development and use of performance information to improve governmental accountability, efficiency, and effectiveness is also likely to be low. Thus, our third hypothesis:

H3: Performance measurement development and outcomes are positively associated with management commitment to the implementation and use of performance information.

2.2.5 Decision-Making Authority

Kwon and Zmud's (1987) review indicates that a second major organizational factor in IT implementation success is the level of worker responsibility. Anderson

Contagion effects can occur when lower-level managers evaluate subordinates using the same criteria used by upper-level managers to evaluate their performance (Hopwood, 1974).

(1995) builds on their definition of worker responsibility to argue that individuals' reactions to management accounting change are positively related to the workers' role involvement, which she defines as "the centrality of the proposed solution to the individuals' jobs, their authority and responsibilities." Consistent with this claim, a subsequent review of ABC implementation studies identifies consistent evidence that implementation success is positively related to the relevance of the information for managers' decisions (Anderson and Young, 1999). These results suggest that managers who believe the innovation can support their decision-making activities are more likely to implement and use the measures. Conversely, managers who lack the authority to make decisions based on the new information will have little reason to embrace the innovation. These results suggest a positive relation between the level of decision-making authority, the extent of system development, and the use of performance information for decision-making.

The hypothesized link between decision-making authority and system implementation and results is also supported by economic theories, which suggest that the level of accountability must be aligned with the decision-rights granted to managers (e.g., Brickley, Smith, and Zimmerman, 1997). This requirement is recognized by government reform advocates, who argue that greater accountability can only be achieved when managers have expanded authority over spending, human resources, and other management functions. As a result, the level of accountability is expected to be positively associated with decision-making authority. However, the requirement for greater authority creates a potential impediment to increased accountability in government organizations, where laws, bureaucratic rules, and the separation of powers

among different branches of government can place severe constraints on managers' decision-making authority, and thereby the extent to which they can be held accountable for results.⁴ Thus, our fourth hypothesis:

H4: Performance measurement development and outcomes are positively associated with the extent to which manager's have the authority to make decisions based on the performance information.

2.2.6 Training

A third organizational factor that is expected to influence the implementation and results of performance measurement innovations is the extent to which resources and training are provided to support the implementation (Kwon and Zmud, 1987; Shields and Young, 1989). Shields (1995) argues that training in the design, implementation, and use of a management accounting innovation allows organizations to articulate the link between the new practices and organizational objectives, provides a mechanism for employees to understand, accept, and feel comfortable with the innovation, and prevents employees from feeling pressured or overwhelmed by the implementation process. The provision of training resources also provides an indication that the organization is providing adequate resources to support the implementation, and signals management support for the innovation (Shields, 1995). If training resources are insufficient, then normal development procedures may not be undertaken, increasing the risk of failure (McGowan and Klammer, 1997).

⁴ The GPRA allows managers to propose, and the Office of Management and Budget to approve, waivers of certain nonstatutory administrative requirements and controls (e.g., procurement authority or greater control over employee compensation). However, the GPRA does not provide agencies with authority to waive requirements for activities within their organizations, and does not allow any waiver of statutory requirements.

Studies of information technology and activity-based costing implementations support these claims, finding positive associations between training investments and implementation success (Kwon and Zmud, 1987; Anderson and Young, 1999).

Accordingly, our fifth hypothesis is:

H5: Performance measurement development and outcomes are positively associated with the extent of related training provided to the manager.

2.2.7 Legislative Mandates

Institutional theory suggests a fourth organizational factor that may be particularly relevant to implementation success in government organizations: whether or not the performance measurement innovation is being implemented in response to legislative mandates or requirements (e.g., Scott, 1987; Covalski and Dirsmith, 1991; Gupta et al., 1994; Brignall and Modell, 2000). Institutional theory argues that organizations gain legitimacy by conforming to external expectations regarding appropriate management control systems in order to appear modern, rational, and efficient to external observers, but tend to separate their internal activities from the externally-focused symbolic systems. In particular, Scott (1987) claims that in institutional environments such as government organizations, where survival depends primarily on the support of external constituents and only secondarily on actual performance, external bodies have the authority to impose organizational practices on subordinate units or to specify conditions for remaining eligible for funding. As a result, subordinate organizations will implement the required practices, but the changes will tend to be superficial and loosely tied to employees' actions.

A number of empirical studies support these theories, finding that government organizations that implement management accounting systems to satisfy legislative requirements make little use of the systems for internal purposes (Ansari and Euske, 1987; Geiger and Ittner, 1996; Brignall and Modell, 2000). Studies of previous management control initiatives in the U.S. government (i.e., Planning, Programming, and Budgeting, Management-by-Objectives, and Zero-Base Budgeting) also indicate that these practices were used more as political strategies for controlling and directing controversy than as tools for improving accountability or decision-making (e.g., Dirsmith et al., 1980). These studies suggest that the recent performance measurement mandates in the U.S. government may increase the development of results-oriented performance measures but have little effect on accountability, use, or performance, leading to our sixth hypothesis:

H6: Performance measurement systems that are implemented to comply with the GPRA's requirements are positively associated with performance measurement development, but are not associated with greater accountability or use of performance data, or with the perceived benefits from GPRA implementation.

2.3 Measurement System Development and System Outcomes

Many government reform advocates contend that the mere availability and reporting of results-oriented performance information fosters improved decision-making by government managers. Consistent with our previous hypotheses, these claims imply a *direct* relation between measurement system development and system outcomes. Others, however, argue that these improvements only occur when the performance measures are used to increase managers' accountability for achieving objectives (e.g., Smith, 1990, 1993; Mayston, 1993; Whynes, 1993; Tirole, 1994; Dixit, 1997), thereby increasing the

managers' incentives to use the information for decision-making. Taken together, these arguments prompt our final hypothesis:

H7: Performance measurement system development has positive *direct* effects on system outcomes, as well as *indirect* effects through the level of accountability for results.

3. Research Design

3.1 Sample

We test our hypotheses using data collected by the United States General Accounting Office (GAO). The GAO survey targeted a random sample of 1,300 middle- and upper-level civilian managers working in the 24 largest executive branch agencies. These agencies represented 97% of the executive branch's full-time workforce and over 99% of the federal government's net outlay in fiscal 1996. The sample was stratified by whether the manager was a member of the Senior Executive Service (SES) and whether the manager worked in an agency or agency component designated as a GPRA pilot.⁵ The questionnaire was pretested using 32 managers from four agencies and revised based on their feedback.

The survey was distributed between November 27, 1996 and January 3, 1997. Managers who did not respond to the initial mailing were sent a follow-up questionnaire. Analysis of responses to the second request revealed no significant differences from

⁵ Members of the Senior Executive Service represent 44.2% of the sample and GPRA pilot sites represent 65.4%. The senior executive stratification was used to control for potential differences in responses by senior managers and lower-level managers by ensuring representative sampling of each group. Stratified sampling of GPRA pilot and non-pilot activities was used because pilot sites were expected to be further along in implementing performance measures than other agencies. The GAO excluded pilots that were designated in fiscal year 1996 because any significant initiatives would have been fairly recent and may not have been sufficiently implemented for any effects to be reflected in questionnaire responses. Most selected pilots were designated in fiscal 1994 and encompassed the entire agency or a major agency component.

earlier responses. Usable surveys were received from 69% of the original sample.⁶ Of the 905 respondents, 108 stated that they did not have performance measures for their activities and are excluded from our tests.⁷

Our initial sample consists of the 797 remaining managers with usable responses. Final sample sizes in our tests range from 380 to 528 due to missing data.⁸

We use the manager of an individual program, project, or operation (henceforth an activity) as our unit of analysis rather than some higher unit (e.g., average responses by all managers within a major program or entire agency) for several reasons. First, many of the survey questions address individual managers' own activities, such as the extent to which respondents have performance measures for the individual programs, projects, or operations they are responsible for, the extent to which they feel accountable for results, and the extent to which they use performance information to manage their activities. Second, field research by the GAO (1997b) finds that the development of performance measures varies significantly *within* a given program or agency, and indicates that

⁶ Of the original sample of 1,300 managers, 47 were eliminated because the individuals had retired, died, left the agency or had some other reason that excluded them from the population of interest, 22 could not be located, 23 refused to participate, 299 questionnaires were not returned, and 4 were returned unusable.

⁷ We exclude managers without performance measures because these managers were not required to answer many of the questions used to develop the constructs used in our analyses. A multivariate logit analysis examining whether a manager had performance measures of any kind found no differences with respect to the type of activity, number of employees, or the percentage of other activities in the same major program that had measures. Senior executives were more likely to have measures for their activities than lower-level managers. Managers with measures also reported greater accountability for achieving results than those without measures. Finally, the presence of performance measures was more likely when the manager belonged to a GPRA pilot site.

⁸ The majority of missing data relates to "no basis to judge" responses to questions. Most of the survey response scales range from 1 = "to no extent" to 5 = "to a very great extent." All of the questions offer a "no basis to judge" response. When this response relates to the respondent's own activities, we code the answer "to no extent," assuming that these topics have little or no impact on an activity if the manager has no basis to respond. In all other cases (e.g., use of performance information for decisions above the respondent's level or perceived results from performance measurement initiatives), "no basis to judge" responses are omitted from the analyses. Final sample sizes for each of the variables used in our tests are provided in Table 1.

managers of some activities have made greater progress implementing measurement systems than others in the same organization. Finally, organizational theory suggests that individual managers are the appropriate unit of analysis because the beliefs and behaviors of individuals toward a particular innovation are shaped by their unique, individual circumstances within the organization (Anderson and Young, 1999).

3.2 Variables

The GAO survey provides substantial information on performance measurement practices and their hypothesized determinants in U.S. government activities. Where possible, we employ multiple indicators for each construct. Factor analysis is used to reduce the dimensionality of the individual questions and minimize measurement error. The resulting multi-indicator constructs are computed using mean standardized responses to the survey questions loading greater than 0.50 on the respective factors. We assess construct reliability for the multi-item variables using factor analysis and Cronbach coefficient alphas. All of the indicator variables pertaining to a given construct load on a single factor, with coefficient alphas above the minimum level suggested by Nunnally (1967) for adequate construct reliability. Specific questions, response scales, and descriptive statistics for the constructs used in our analyses are provided in Table 1.

3.2.1 Measurement System Development

System development is assessed using the variable MEASUREMENT, which captures the extent to which respondents have developed different types of results-oriented performance measures (where 1 = to no extent and 5 = to a very great extent) for the activities they are involved with, from the following list: quantity of products or services, operating efficiency, customer satisfaction, product or service quality, and

measures that demonstrate to someone outside the agency whether the organization is achieving its intended results.⁹

3.2.2 *System Outcomes*

We evaluate system outcomes using three constructs capturing the stated objectives of governmental performance measurement efforts: greater accountability, enhanced decision-making, and improved governmental performance.¹⁰

Four questions measure the extent to which managers feel they are held *accountable* for results. Respondents were asked to rate the following statements on a five-point scale (where 1 = to no extent and 5 = to a very great extent): (1) managers at my level are accountable for the results of the program(s)/project(s)/operations(s) they are responsible for, (2) employees in my agency receive positive recognition for helping the agency accomplish its strategic goals, (3) the individual I report to periodically reviews with me the results or outcomes of the program(s)/project(s)/operations(s) I am responsible for, and (4) the *lack* of incentives (e.g., rewards or positive recognition) has *hindered* using performance information. The last question is reverse-coded when developing the construct.

⁹ The fact that all of the performance measure categories load on a single factor indicates that managers of activities tend to implement all of these measures together. This is consistent with theories calling for greater measurement diversity in strategic performance measurement systems, but is inconsistent with theories stating that the types of measures should be tailored to reflect the organization's strategies or the specific actions desired of agents in multitasking environments. See Ittner et al. (2002) for a discussion of these theories. Additional analysis by type of activity and other contingency variables provided no additional insight into the greater combined use of all these variables. However, the performance measurement categories in the survey are consistent with the GPRA's requirements for output, service level, and outcome measures for each activity. Consequently, the greater implementation of measures related to each of these categories may reflect efforts to meet the Act's requirements.

¹⁰ Our outcome variables are similar to those used to evaluate the success of activity-based costing implementations. See, for example, Foster and Swenson (1997) and Anderson and Young (1999).

Eleven questions address the *use* of performance measures. Factor analysis with oblique rotation indicates that these questions represent two underlying constructs. Eight questions loading greater than 0.50 on the first factor reflect lower-level uses related to the managers' own activities (denoted MGR USE). These questions ask the extent to which respondents use performance information for the activities they are involved with when: (1) setting program priorities, (2) allocating resources, (3) adopting new program approaches or changing work processes, (4) coordinating program efforts with other internal or external organizations, (5) refining program performance measures, (6) setting new or revising existing performance goals, (7) setting individual job expectations for subordinates, and (8) rewarding subordinate government employees.

Three questions loading greater than 0.50 on the second factor emphasize higher-level uses of performance information (denoted HIGHER USE). These questions address the extent to which performance information is used to develop the agency's budget, make funding decisions, and make management changes above the respondent's organizational level.

Finally, we examine the benefits from the U.S. government's recent performance measurement mandates using two questions on the *perceived results* from the Government Performance and Results Act. While government reform advocates contend that the GPRA's externally-imposed reporting practices will improve governmental performance (particularly in the presence of greater accountability), institutional theory argues that *mandated* practices will have little effect on governmental performance regardless of the extent of system implementation. The two questions ask the extent to which respondents believe that efforts to implement the GPRA have improved their

organizations' activities to date (denoted RESULTS TO DATE), or will improve them in the future (denoted FUTURE RESULTS). Since many respondents were not sufficiently involved in GPRA efforts to have an opinion on its current effects, we treat each question separately.

3.2.3 Implementation Factors

Following Kwon and Zmud (1987), Shields and Young (1989), and others, we examine both technical and organizational influences on the measurement system outcome variables. The variables used to measure the hypothesized implementation factors are discussed below.

3.2.3.1 Data Limitations and Metric Difficulties

The survey contains 11 questions on potential factors hindering performance measurement and management. Consistent with discussions in the performance measurement literature, factor analysis with oblique rotation reveals two underlying dimensions with eigenvalues greater than one.¹¹ Four questions loading greater than 0.50 on the first factor (denoted DATA LIMITATIONS) emphasize limitations in existing information systems' ability to provide required data. These questions address difficulties obtaining valid or reliable data, difficulties obtaining data in time to be useful, the high cost of collecting data, and the inability of existing information systems to provide the needed data.

Five questions loading greater than 0.50 on the second factor (denoted METRIC DIFFICULTIES) relate to problems defining and interpreting performance metrics. The

¹¹ Questions concerning implementation problems were only asked to respondents who had performance measures for their activities. Two questions about (1) different parties using different definitions to measure performance, and (2) difficulty resolving conflicting interests of internal and/or external stakeholders did not load 0.50 or greater on any factor. These questions are not included in our analyses.

questions ask managers the extent to which they have experienced difficulties determining meaningful measures, associating their activities with future results, distinguishing results due to their activities from other factors, and determining how to use performance information to improve activities or set goals.

3.2.3.2 Management Commitment

We develop the construct COMMITMENT to measure the extent to which top leadership is committed to achieving results via performance measurement. COMMITMENT is based on three questions: (1) to what extent does the agency's top leadership demonstrate a strong commitment to achieving results, (2) to what extent has the *lack* of ongoing top executive commitment to using performance information to make program/funding decisions *hindered* measuring performance or using performance information, and (3) to what extent has the *lack* of ongoing congressional commitment to using performance information to make program/funding decisions *hindered* measuring performance or using performance information. The latter two questions are reverse-coded when computing the construct.

3.2.3.3 Decision-Making Authority

The level of decision-making authority (denoted AUTHORITY) is assessed using responses to a single question asking whether managers at the respondent's level have the decision-making authority they need to help the agency accomplish its strategic goals.

3.2.3.4 Training

Respondents were asked whether they have received training to accomplish the following measurement-related tasks: (1) conduct strategic planning, (2) set program performance goals, (3) develop program performance measures, (4) use program

performance information to make decisions, and (5) link the performance of program(s)/operation(s)/project(s) to the achievement of agency strategic goals. We code each response one if the agency provided training in that task, and zero otherwise. The construct TRAINING represents the sum of the individual responses.

3.2.3.5 Legislative Mandates

We proxy for the effects of legislative mandates on performance measurement implementation using an indicator variable for GPRA pilot sites. The GAO (1997b) argues that pilot sites are likely to have more highly developed measurement systems than other sites due to their earlier efforts to meet the GPRA's mandate for results-oriented performance measures. However, the GAO makes no assessment of whether this information is actually used to improve accountability or decision-making. The variable PILOT is coded one if the activity was part of a GPRA pilot, and zero otherwise.

3.2.4 Control Variables

We include two control variables in our tests. Our first control is an indicator variable for members of the Senior Executive Service (denoted SES). This variable is included to control for potential differences in responses between senior and lower-level managers. We also include a second control variable in models examining perceived GPRA benefits to account for potential biases in responses by those participating in the implementation process. GPRA INVOLVEMENT represents the average standardized response to two questions on the involvement of managers and their staff in GPRA implementation efforts.¹²

¹² To examine the robustness of our results to model specification, we repeated the analyses using a number of other control variables, including the natural logarithm of the number of employees in the activity (a size control), the type of activity managed by the respondent (internal agency efforts, federal government-wide support, research and development, service delivery, and other), and a program control for organizational

4. Descriptive Statistics

Descriptive statistics are provided in Table 1.¹³ The most highly-developed measures are volume indicators, with 60.8% of managers having these measures to a great or very great extent. The least developed measures relate to operating efficiency, with only 44.7% of managers having these measures to a great or very great extent.

Almost 60% of respondents feel that managers at their level are held accountable for results to a great or very great extent. However, fewer than half (47.6%) note that their superior extensively reviews their results with them on a periodic basis. Less than a quarter believe that the lack of incentives has severely hindered using performance information.

Between 59.6% and 68.8% of the respondents report using performance measures extensively for managerial purposes, depending upon the type of measure. There is considerably lower perceived use of performance measures for higher-level decisions. Only 28.9% believe that results-oriented performance information has a major influence on budgets, the most *extensive* higher-level use. The least common use of performance information is for program, operation, or project changes by upper-level management, with only 23.1% of managers believing that upper-level management extensively uses the performance information for these purposes.

effects on the managers' responses (measured using the average response by other managers in the same program). These controls had virtually no effect on our results and are excluded from the reported models.

¹³ Although average standardized responses are used to compute some of the constructs, we report unstandardized responses in Table 1 to provide insight into the performance measurement practices in our sample. Means (standard deviations) for the standardized constructs are -0.002 (0.182) for MEASUREMENT, 0.048 (0.700) for ACCOUNTABILITY, 0.005 (0.830) for MGR USE, 0.100 (0.873) for HIGHER USE, 0.006 (0.821) for DATA LIMITATIONS, 0.007 (0.753) for METRIC DIFFICULTIES, 0.021 (0.764) for COMMITMENT, and 0.461 (0.498) for GPRA INVOLVEMENT.

Most managers rate the benefits from GPRA implementation relatively low. Only 13.7% feel that the GPRA has improved agency performance to a great or very great extent to date, with 34.7% feeling it will have a great or very great impact in the future. In contrast, 52.3% believe the GPRA has had little or no impact to date, while 29.9% believe its impact will be small to nonexistent in the future (not shown in the table).

4.1 Correlations

Table 2 provides Spearman correlations among the variables used in our study. More than 75% of the associations are significant at the 5% level or better (two-tailed).¹⁴ Performance measure development, accountability, and uses are positively related to each other, negatively related to system and measurement problems, and positively related to the extent of management commitment, decision-making authority, and training. These variables are also positively related to whether the manager is a senior executive (SES) and the extent of GPRA involvement.

The perceived benefits of GPRA-related activities (both to date and in the future) are positively associated with performance measure development, accountability, and use. Organizations that demonstrate a strong commitment to results are also more likely to allow greater decision-making authority, to provide more training, to have a greater proportion of senior executive respondents, and to have greater GPRA involvement.

¹⁴ Pearson correlations are virtually identical and are available from the authors upon request. Despite the significant correlations, all Variance Inflation Factor (VIF) scores are below 2.5, indicating no serious problems with multicollinearity in subsequent regression models.

5. Results

5.1 Performance Measure Development

Table 3 provides evidence on the determinants of results-oriented performance measure development. Due to missing responses for some of the variables, the sample size is 528 in this analysis. The resulting regression is highly significant, with an adjusted R^2 of 30%.

Most of the results support our hypotheses.¹⁵ Metric difficulties (i.e., difficulties determining meaningful measures, results occurring too far into the future to be measured, difficulties distinguishing between results produced by the program and results caused by other factors, and difficulties determining how to use performance information to improve the program or to set new or revise existing performance goals) significantly dampen the extent of performance measure development. Top management commitment, decision-making authority, and the level of training provided to managers all exhibit significant positive associations with performance measure development. Moreover, GPRA pilot sites have performance measures to a greater extent than non-pilots, indicating that efforts to meet the Act's requirements have increased measurement system development.

One result that differs from our hypotheses is the insignificant relation between data limitations (i.e., difficulties obtaining valid or reliable data, difficulties obtaining data in time to be useful, and the high cost of data collection) and the development of performance measurement systems. Contrary to Hypothesis H1, data limitations do not appear to affect measurement system development. The coefficient on SES is also

¹⁵ One-tailed tests are used for all of the variables with hypothesized signs and two-tailed tests are used for control variables. Variables with p-values of 0.05 or less are considered statistically significant.

statistically insignificant, indicating that measurement system development is no higher for senior executives' activities than for lower-level activities.

One limitation to the preceding analysis is the assumption that the various technical and organizational factors independently influence the extent of performance measurement development. However, it is possible that these factors interact to impact the development of results-oriented performance measures. Given the large number of potential interactions and limited theory on how these factors interrelate, we employ an exploratory technique called CHAID (CHi-squared Automatic Interaction Detection) to examine whether interactions among the predictor variables have significant effects on measurement development. CHAID modeling selects a set of predictors and their interactions that optimally predict the dependent variable. The technique assesses whether splitting the sample based on the predictor variables leads to a statistically significant discrimination ($p < 0.05$) in the dependent variable using either chi-squared tests or F-tests, depending upon whether the predictor variable is categorical or continuous. The first split represents the difference in a single predictor variable that is most significant in explaining differences in the dependent variable. This splitting continues until no further split of a predictor variable provides significant differences in the dependent variable. The final splits, or "terminal nodes", represent subgroups of observations that are maximally different from each other on the dependent variable, and can be characterized by the scores for the various predictor variables used to split the sample into these subgroups.¹⁶

¹⁶ Another advantage of CHAID analysis is the ability to detect non-linearities in the associations between the predictor variables and the dependent variable. See Breiman (1984) and *AnswerTree* (1998) for discussions of CHAID and other related methods.

...

The CHAID analysis (not reported in the tables but available from the authors) indicates that the highest MEASUREMENT scores are found in government activities that have received training in all five measurement-related topics and have relatively low metric difficulties and data limitations (mean standardized MEASUREMENT score = 0.88). The lowest MEASUREMENT scores are found in activities with training in fewer than five of the measurement-related topics, extensive problems selecting and interpreting appropriate performance metrics, and low decision-making authority (mean = -0.72). High management commitment increases the development of performance measures in activities that have received training in fewer than five of the surveyed topics and have medium levels of metric difficulties (mean = 0.43 in these activities vs. -0.05 in activities with similar training levels but relatively low management commitment), but is not a significant discriminator of performance measure development in the other subgroups.

Overall, the preceding findings are consistent with prior studies on the *organizational* determinants of information system and management accounting innovation. However, our findings regarding the *technical* problems associated with the development of organizational performance measures are mixed. We find no evidence that data limitations are related to performance measure development in the regression models. Moreover, the CHAID analyses indicate that data limitations only influence measurement system development when training is extensive and metric difficulties are low, in which case fewer data limitations are associated with greater measurement system development. In contrast, we find strong and consistent evidence that difficulties selecting and interpreting metrics have a negative impact on performance measurement

implementation. These results suggest that problems identifying appropriate measures and understanding their causal relationships will be particularly important as more public and private sector organizations attempt to implement systems to measure “intangible assets” and “intellectual capital,” and to develop organizational models of leading and lagging indicators of performance.

5.2 Accountability

We next examine factors influencing the outcomes from measurement system development. As Smith (1990) notes, one of the keys to evaluating the effectiveness of a governmental information system is determining the extent to which the system allows principals (i.e., citizens and their elected or appointed representatives) to satisfactorily control their agents. Evidence on the determinants of accountability is presented in Table 4. The model is highly significant, and explains 51% of the variation in the ACCOUNTABILITY construct.

Perhaps the most important question is whether performance measure development is associated with increased accountability, as emphasized in the government reform literature. Consistent with Hypothesis H7, the extent of performance measure development is positively associated with the extent to which government managers are held accountable for results ($p < 0.001$). The positive and significant association supports claims that the reporting of governmental performance information enhances principals’ ability to hold their agents accountable for results.¹⁷

¹⁷ Following discussions in the government reform literature, the tests in Table 4 assume that the extent of accountability is a function of performance measure development. However, prior studies suggest that the direction of causality may run from incentives to system development since employees need to see the link between incentives and the system innovation to support its implementation (e.g., Shields, 1995; Anderson and Young, 1999). Moreover, economic theories suggest that accountability and performance measurement levels should be simultaneously determined (e.g., Rose-Ackerman, 1986). To examine the direction of causality, we estimated a simultaneous equations model with ACCOUNTABILITY and

Several of the implementation factors also explain differences in accountability, even after controlling for their influence on measurement development. These results suggest that some of the implementation factors have direct effects on accountability, as well as indirect effects through measurement development. Management commitment, decision-making authority, and training in performance measurement topics all exhibit significant, positive direct and indirect effects on accountability. Thus, the extent to which government managers are held accountable for achieving results is influenced not only by the extent of performance measurement, but also by managers' knowledge of and ability to apply results-oriented management techniques and by top management's commitment to achieving results.¹⁸

In contrast to the insignificant relation with performance measure development, data limitations (i.e., problems providing necessary, relevant, and valid performance data in a timely and cost effective manner) are *negatively* associated with the extent to which managers are held accountable for results. Difficulties selecting and interpreting appropriate performance metrics (METRIC DIFFICULTIES), on the other hand, have no direct effect on accountability, even though they are significantly associated with the extent of measurement. These results provide mixed support for our hypotheses, and

MEASUREMENT as dependent variables. Following the results in Tables 4 and 5, DATA LIMITATIONS served as the instrument for ACCOUNTABILITY and METRIC DIFFICULTIES as the instrument for DEVELOPMENT. The coefficient on MEASUREMENT was positive and significant in the ACCOUNTABILITY model ($p < 0.023$, one-tailed), but ACCOUNTABILITY was not significant in the DEVELOPMENT model ($p = 0.35$, one-tailed). Thus, the extent of accountability appears to be a function of performance measure development in this setting.

¹⁸ The positive association between top management commitment and accountability is not surprising since it is unlikely that managers who are not committed to the use of performance information would hold their subordinates accountable for achieving performance objectives. However, the model's significant explanatory power is not primarily due to this association. When COMMITMENT is removed from the model, the adjusted R^2 falls from 0.51 to 0.44, indicating that the management commitment variable only explains approximately 7% of the variation in the accountability construct.

suggest that data and metric problems have different effects on performance measurement outcomes. Problems developing performance metrics appear to be a significant impediment to the initial development of performance measurement systems, but to have little influence on the use of the resulting system for holding managers accountable once these problems are resolved. In contrast, data problems do not impede the development of the measurement system, but tend to deter government officials from using the resulting system for performance evaluation. The latter result is consistent with Krumwiede's (1998) finding that information system issues have a significant influence on whether organizations undertake later stages of ABC adoption, but not on whether they undertake earlier stages.

Despite the previous evidence that GPRA pilot sites have developed performance measures to a greater extent than non-pilot sites, we find no evidence that pilot sites hold managers accountable for results to a greater extent than other units. This finding supports institutional theories that government organizations implement management control systems to meet legislative requirements but do not use these systems for internal purposes (Hypothesis H6).¹⁹ Additionally, senior executives (SES) feel *less* accountable for results than do lower-level government managers.

We again use exploratory CHAID analysis to examine potential interactions among the hypothesized determinants of accountability. The CHAID results (not reported in the tables) indicate that the most important predictors of

¹⁹ Further support for this conclusion is provided by our analysis of performance measurement *changes* over the past three years (not reported in the tables). The survey provided data on both current and past performance measurement and accountability practices. The GAO argues that most of these recent changes have been prompted by new government requirements for performance information. Although reported measurement levels are statistically larger than those three years prior ($p < 0.05$, two-tailed t-test), these increases are not statistically associated with changes in accountability, again suggesting that recent performance measurement mandates are not achieving their goal of promoting greater accountability.

ACCOUNTABILITY are the level of decision-making authority and the extent of management commitment. Activities reporting “very extensive” authority have the highest mean standardized ACCOUNTABILITY scores (1.05), while activities reporting no decision-making authority have the lowest mean scores (-0.81). For the other three levels of decision-making authority, the extent of accountability depends upon management’s commitment to the use of performance information. For example, in activities reporting “extensive” decision-making authority, the mean ACCOUNTABILITY score equals 0.78 in the subgroup with very high management commitment, 0.55 in those with relatively high commitment, 0.31 when commitment is relatively low, and -0.01 when management commitment is very low. Similar results are found in activities with decision-making authority scores of 2 or 3 (where 1 = “to no extent” and 5 = “to a very great extent”). Thus, the interaction between the level of decision-making authority and the extent of management commitment to the use of performance information has a significant impact on the level of accountability for results only when decision-making authority is neither very high nor very low in activities where decision-making authority is neither extremely high nor extremely low.

5.3 Use of Performance Information

Table 5 investigates the factors influencing the use of performance information for lower-level and higher-level decision-making. The tests provide strong evidence that the extent of performance measure development and accountability are positively related to the use of results-oriented performance information ($p < 0.001$), both by managers for their own activities and for higher-level decisions. Together with the results in Table 4, this evidence suggests that greater performance measure development has both direct

effects and indirect effects (through accountability) on the use of performance information. These results are consistent with Hypothesis H7, and support claims that the benefits from the development of measurement systems are greater when managers are held accountable for results. However, the significant direct effects are inconsistent with claims that managers *must* be held accountable for governmental performance measurement initiatives to be effective.

In contrast to Hypothesis H1, data limitations are *positively* associated with the use of performance information at the manager's level and for higher-level decisions. One explanation for these results is that managers do not experience significant problems with information systems and data collection until the information is actually being used for decision-making. This interpretation is consistent with our earlier findings that data limitations do not prevent performance measures from being developed, but do make it more difficult to hold managers accountable for results.

As predicted, difficulties selecting and interpreting performance metrics are negatively associated with lower-level managerial uses of performance information. However, these difficulties are not directly associated with higher-level uses after controlling for system development and accountability. In contrast, top management commitment, decision-making authority, and training are all positively associated with greater higher-level uses of performance information, but not lower-level uses. These differential results suggest that any effects of management commitment, decision-making authority, and training on the respondents' use of performance information for managing their own activities come *indirectly* through the influence of greater measurement system development and perceived accountability for results. However, these implementation

factors also appear to have both direct and indirect effects on respondents' beliefs about the use of performance information by superiors. Once again, the coefficient on the GPRA pilot indicator variable is insignificant, supporting the hypothesis that greater measurement system development in response to the Act's requirements has not translated into greater use of the information for internal purposes.

CHAID analysis of interactive effects (not reported in the tables) indicates that the level of performance measure development and the extent of metric difficulties are the most significant determinants of the use of performance information for lower-level decisions. MGR USE progressively increases as the level of performance measure development increases (mean = -0.72 in the subgroup with the lowest MEASUREMENT scores, -0.07 in the next lowest subgroup, 0.24 in the subgroup with moderately high MEASUREMENT scores, and 0.65 in the high MEASUREMENT score subgroup). Moreover, when MEASUREMENT is moderately high, fewer problems selecting and interpreting performance metrics are associated with higher managerial use of performance information (-0.11 when metric difficulties are high, 0.22 when metric difficulties are moderate, and 0.43 when metric difficulties are low). No other interaction is significant in the CHAID analysis. These results again suggest that greater availability of performance measures leads to greater use of this information for decision-making, but indicate that implementing relatively extensive performance measurement system, without overcoming problems selecting and interpreting appropriate performance metrics is likely to have little effect on managers' actions.

The CHAID analyses also identify a number of interactive effects on higher-level uses of performance information. The largest HIGHER USE scores are found in activities

reporting “very extensive” decision-making authority (mean standardized HIGHER USE score = 0.87). This is followed by activities reporting “extensive” authority *and* high levels of measurement system development (mean = 0.72). If the manager of the activity reported “extensive” decision-making authority but relatively low measurement system development, the mean HIGHER USE score falls to 0.02. Similarly, in activities with medium levels of authority (3 on the 1 to 5 scale), HIGHER USE has a score of 0.44 when MEASUREMENT is high, -0.002 when MEASUREMENT is medium, and -0.36 when MEASUREMENT is low. The lowest usage scores are found in activities with low decision-making authority *and* low management commitment to the use of performance information (mean = -0.95). Limited higher-level usage is also found in activities with low decision-making authority, somewhat higher levels of management commitment, but low levels of measurement system development (mean = -0.47). This evidence suggests that decision-making authority, measurement system development, and management commitment have complex, non-linear interactive effects on the use of performance information for higher-level decision.

5.4 Perceived GPRA Benefits

Our final tests examine the influence of the hypothesized implementation factors and accountability on the perceived benefits from implementing the GPRA’s *mandated* requirements. Table 6 displays results on the perceived benefits to date and in the future. The models regress perceived benefits on the predictor variables used in our earlier analyses and the extent of the manager’s participation in GPRA implementation efforts (GPRA INVOLVEMENT).²⁰

²⁰ Due to missing responses, the sample size is 380 when results to date is the dependent variable, and 434 when expected future results is the dependent variable.

We find mixed evidence that performance measure development and accountability are related to the perceived benefits from implementing the GPRA's requirements. Performance measure development is positively related to perceived results to date, but is unrelated to expected results in the future. Furthermore, accountability is unrelated to results to date, and *negatively* related to expected future results. At best, these results provide only weak support for the claimed benefits from *mandated* increases in performance measurement and accountability in government organizations. The limited perceived benefits are consistent with institutional theories that the implementation of mandated organizational changes in government organizations tends to be symbolic, with little effect on internal operations.

The estimated coefficients on the other predictor variables also provide mixed support for our hypotheses. Consistent with the usage results, data limitations are *positively* associated with perceived benefits to date and in the future. One potential explanation for the significant positive coefficients on DATA LIMITATIONS is that managers who have encountered impediments such as poor information systems and the high cost of data collection believe that implementation of the GPRA's requirements has helped and will continue to help overcome these problems and improve performance.²¹ Difficulties selecting and interpreting appropriate performance metrics (METRIC DIFFICULTIES), on the other hand, are *negatively* associated with the expected future benefits from the U.S. government's performance measurement initiatives, suggesting

²¹ The positive associations between data problems and GPRA results are not driven by the positive relations between data limitations and the use of performance information identified in Table 5. When the two usage variables are included in the GPRA results models, the coefficients on DATA LIMITATIONS remain positive and significant, while the significance levels of the other coefficients change little.

that managers believe these problems will be difficult to overcome even with the GPRA initiatives.

Providing managers with the decision-making authority they need to help the agency accomplish its strategic goals is positively and significantly associated with perceived benefits to date and in the future. Together with the earlier results, this evidence provides strong support for claims that managers need decision-making authority to achieve significant benefits from performance measurement innovations.

Managers who receive more extensive training in measurement-related topics are more likely to believe the GPRA is or will be beneficial. Likewise, managers who are more actively involved in the GPRA's implementation rate the Act's potential benefits higher than managers with little involvement.

Strong commitment on the part of top leadership, on the other hand, is unrelated to the perceived benefits from the GPRA to date or in the future, despite the generally significant associations between commitment and measurement accountability and use. Senior executives also perceive the future benefits from implementing the GPRA's requirements to be lower than do lower-level managers. Despite the more extensive development of performance measures in GPRA pilot sites, managers of these activities do not rate the benefits from fulfilling the GPRA's requirements any higher than do managers of non-pilot activities. The insignificant coefficients again suggest that extensive implementation of the GPRA's mandated requirements has little influence on internal operations.

CHAID analyses (not reported in the tables) suggest that interactive effects exist among some of the the predictor variables. With perceived results to date, the highest

scores are found in activities with *medium* levels of measurement system development, extensive training, and high management commitment (mean = 3.4 on a scale from 1 = to no extent and 5 = to a very great extent). Not surprisingly, the lowest perceived results are in activities that have undertaken little measurement system development (mean = 1.43). For perceived future results, the highest scores are found in activities with few data limitations and *medium* levels of training (mean = 3.41). The lowest scores, in turn, are found in activities with extensive data limitations and *very high* levels of measurement system development (mean = 2.49). The latter finding is inconsistent with the regression results in Table 6, but again suggests that data limitations do not become a serious problem until the performance measurement system is extensively developed.

6. Conclusions

This study draws upon the information systems change, management accounting innovation, and public sector reform literatures to examine some of the factors influencing the implementation, use, and perceived benefits of results-oriented performance measurement systems in the U.S. government. We find that performance measure development and accountability are hindered by factors such as inadequate training, the inability of existing information system to provide timely, reliable, and valid data in a cost effective manner, difficulties selecting and interpreting appropriate performance measures, lack of organizational commitment to achieving results, and limited decision-making authority. These issues are likely to be equally important in the private sector as firms implement strategic performance measurement systems that capture less-traditional performance information.

We also find that GPRA pilot sites have developed performance measures to a greater extent to meet the Act's requirements, but do not make greater use of the information. This result is consistent with institutional theories, which contend that implementation of externally-mandated control systems is likely to be symbolic, with little influence on internal operations. In contrast, increased performance measurement development and accountability are positively associated with the use of performance information after controlling for GPRA implementation efforts, supporting claims that internal performance measurement efforts and greater accountability for results can provide the necessary information and incentives for performance-based management, even in the absence of mandates. Although greater measurement and accountability are positively associated with the use of performance information for decision-making, we find only weak evidence that performance measure development and increased accountability influence managers' perceptions of the benefits from complying with the GPRA's reporting mandates, contradicting the assumptions underlying most initiatives to improve governmental performance through *mandated* reporting requirements. Finally, our exploratory CHAID analyses indicate that some of these technical and organizational factors can have (sometimes complex) interactive effects on performance measurement system implementation and outcomes.

The findings from this study are not without limitations. First, we are limited to perceptual measures, rather than "hard" measures such as the actual number and frequency of performance measures or actual outcomes. Although the perceptual measures are similar to those used in other survey-based management accounting studies, future investigations can make a significant contribution by examining the actual outcomes

associated with the implementation of results-oriented measurement systems. Second, we do not have data on a number of potential factors associated with performance measurement or the use of performance information, such as the activity's competitive environment and the type or source of funding received by the organization (e.g., Geiger and Ittner, 1996; Brignall and Modell, 2000). Third, the survey did not provide information on the target-setting process or the level of target achievability, which are likely to have a significant impact on the benefits from performance measurement initiatives. Finally, the surveyed measurement systems may not have been in place long enough to provide a true reflection of their benefits. Although most of the GPRA pilot sites began implementing their systems more than three years prior to the study, many of the organizations may not have had enough time to integrate the new systems into their day-to-day activities. However, a more recent GAO (2001) survey indicates that many of the implementation issues identified in our study, such as the lack of top management commitment and limited decision-making authority, remain common in the U.S. government. Further analysis of the GPRA's external reporting requirements provides a natural opportunity for researchers to examine the maturation in performance measurement and management control practices and the ongoing performance gains from their use.

References

- Anderson, S.W.. 1995. A framework for assessing cost management system change: the case of activity-based costing implementation at General Motors, 1986-1993. *Journal of Management Accounting Research*, 7, 1-51.
- Anderson, S.W. and S.M. Young. 1999. The impact of contextual and process factors on the evaluation of activity-based costing systems. *Accounting, Organizations and Society*, 24, 525-559.
- Ansari, S.L., and K.J. Euske. 1987. Rational, rationalizing, and reifying uses of accounting data in organizations. *Accounting, Organizations and Society*: 549-579.
- AnswerTree 1.0 User's Guide*, Chicago, IL: SPSS Inc., 1998.
- Atkinson, Anthony A. and James Q. McCrindell. 1997. Strategic performance measurement in government. *Cost & Management*: 20-23.
- Atkinson, Anthony A., John H. Waterhouse, and Robert B. Wells. 1997. A stakeholder approach to strategic performance measurement. *Sloan Management Review*: 25-37.
- Breiman, L., 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.
- Brickley, J., Smith, C., Zimmerman, J., 1997. *Managerial Economics and Organizational Architecture*. Richard D. Irwin, Burr Ridge, IL.
- Brignall, Stan, and Sven Modell. 2000. An institutional perspective on performance measurement and management in the 'new public sector'. *Management Accounting Research* 11: 281-306.
- Covaleski, Mark, and Mark Dirsmith. 1991. The management of legitimacy and politics in public sector accounting. *Journal of Accounting and Public Policy* 10 (Summer): 135-157.
- Dirsmith, Mark W., Stephen F. Jablonsky, and Andrew D. Luzi. 1980. Planning and control in the U.S. government: a critical analysis of PPB, MBO and ZBB. *Strategic Management Journal* 1: 303-329.
- Dixit, Avinash. 1997. Power of incentives in private versus public organizations. *American Economic Review* 87(2) (May): 378-382.
- Doll, W.J. 1985. Avenues for top management involvement in successful MIS development. *MIS Quarterly* 9: 17-35.

- Flynn, Norman. 1986. Performance measurement in public sector services. *Policy and Politics* 14(3): 389-404.
- Foster, G. and D.W. Swenson. 1997. Measuring the success of activity-based cost management and its determinants. *Journal of Management Accounting Research* 9 (Fall): 109-141.
- GAO. 1997a. Managing for results: analytic challenges in measuring performance. Gaithersburg, MD: General Accounting Office, GAO/HEHS/GGD-97-138, May.
- GAO. 1997b. The government and performance results act: government-wide implementation will be uneven. Gaithersburg, MD: General Accounting Office, GAO/GGD-97-109, June.
- GAO. 2001. Managing for results: federal managers' views on key management issues vary widely across agencies. Gaithersburg, MD: General Accounting Office, GAO-01-592, May.
- Gates, S., 1999. *Aligning Strategic Performance Measures and Results*. The Conference Board, New York, NY.
- Geiger, Dale R. and Christopher D. Ittner. 1996. The influence of funding source and legislative requirements on government cost accounting practices. *Accounting, Organizations and Society* 21 (August): 549-567.
- Gupta, Parveen P., Mark W. Dirsmith, and Timothy J. Fogarty. 1994. Coordination and control in a government agency: institutional theory perspectives on GAO audits. *Administrative Science Quarterly* (June): 264-284.
- Holmstrom, B. and P. Milgrom. 1991. Multitask principal-agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics, and Organization* 7: 24-52.
- Hood, C. 1995. The "new public management" in the 1980s: variations on a theme. *Accounting, Organizations and Society*: 93-109.
- Hopwood, A.G. 1974. Leadership climate and the use of accounting data in performance evaluation. *The Accounting Review*: 485-495.
- Ittner, C.D. and D.F. Larcker. 1998. Innovations in performance measurement: trends and research implications. *Journal of Management Accounting Research*: 205-238.
- Ittner, C.D., D.F. Larcker, and T. Randall. 2002. Performance implications of strategic performance measurement in financial services firms. Working paper, University of Pennsylvania and University of Utah.

- Jones, L.R. 1993. Counterpoint essay: nine reasons why the CFO act may not achieve its objectives. *Public Budgeting & Finance* 13(1): 87-94.
- Jones, L.R. and Jerry L. McCaffery. 1997. Implementing the chief financial officers act and the government performance and results act in the federal government. *Public Budgeting & Finance* 17(1): 35-55.
- Kravchuk, Robert S. and Ronald W. Schack. 1996. Designing effective performance-measurement systems under the government performance and results act of 1993. *Public Administration Review* 56 (4): 348-358.
- Krumwiede, K., 1998. The implementation stages of activity-based costing and the impact of contextual and organizational factors. *Journal of Management Accounting Research* 10, 239-277.
- Kwon, T. and Zmud, R. 1987. Unifying the fragmented models of informations systems implementation. In R.J. Boland and R. Hirschiem, *Critical issues in informations systems research*. New York: John Wiley.
- Macintosh, N.B. 1985. *The social software of accounting and information systems*. New York, NY: John Wiley & Sons.
- Manley, J.H. 1975. Implementation attitudes: A model and measurement methodology. In R.L. Schultz and D.P. Slevin, *Implementing operations research/management science*, New York, NY: American Elsevier: 183-202.
- Mayston, David. 1993. Principals, agents and the economics of accountability in the new public sector. *Accounting, Auditing & Accountability Journal* 6(3): 68-96.
- McGowan, Annie S. and Thomas P. Klammer. 1997. Satisfaction with activity-based cost management implementation. *Journal of Management Accounting Research* 9: 217-237.
- Nunnally, J., 1967. *Psychometric Theory*, McGraw-Hill, New York, NY.
- Osborne, D. and T. Gaebler. 1993. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. New York: Addison-Wesley.
- Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 38: 7-63.
- Rose-Ackerman, Susan. 1986. Reforming public bureaucracy through economic incentives? *Journal of Law, Economics, and Organization* 2: 131-161.

- Schultz, R. and Ginzberg, M.J. 1984. Implementation research—the third generation. *Applications of Management Science* 4: 1-83.
- Scott, W.R. 1987. The adolescence of institutional theory. *Administrative Science Quarterly* 32 (4): 493-511.
- Shields, M., 1995. An empirical analysis of firm's implementation experiences with activity-based costing. *Journal of Management Accounting Research* 7, 1-28.
- Shields, M. and Young, S.M. 1989. A behavioral model for implementing cost management systems. *Journal of Cost Management* 2, 29-34.
- Smith, Peter. 1990. The use of performance indicators in the public sector. *Journal of the Royal Statistical Society A* 153 (Part 1): 53-72.
- Smith, Peter. 1993. Outcome-related performance indicators and organizational control in the public sector. *British Journal of Management* 4: 135-151.
- Tirole, Jean. 1994. The internal organization of government. *Oxford Economic Papers* 46(1) (January): 1-29.
- U.S. Senate. 1992. Government performance and results act. Washington, D.C.: Committee on Governmental Affairs, Report 102-429, September 29.
- Whynes, David K. 1993. Can performance monitoring solve the public services' principal-agent problem? *Scottish Journal of Political Economy* 40(4) (November): 434-446.

Table 1

Summary Statistics for the Survey Questions Used to Develop the Measurement System Development, System Outcome, and Implementation Factor Variables

Construct and Survey Items	Mean	Std. Dev.	%Great or Very Great Extent^b
MEASUREMENT (n = 757; coefficient $\alpha = 0.87$)			
To what extent do you have the following performance measures for your activities? ^a			
1. Quantity of products or services provided	3.63	1.15	60.8%
2. Operating efficiency	3.25	1.16	44.7%
3. Customer satisfaction	3.22	1.20	45.2%
4. Quality of products or services provided	3.25	1.16	46.6%
5. Measures demonstrating to external parties whether or not you are achieving intended results	3.36	1.14	51.2%
ACCOUNTABILITY (n = 744; coefficient $\alpha = 0.70$)			
To what extent do you agree with the following statements? ^a			
1. Managers at my level are held accountable for the results of their activities	3.59	1.02	59.8%
2. Employees in my agency receive positive recognition for helping the agency accomplish strategic goals	3.07	1.05	36.1%
3. The individual I report to periodically reviews my activity's results with me	3.26	1.20	47.6%
4. Lack of incentives (e.g., rewards, positive recognition) has <u>hindered</u> using performance information (reverse coded in the construct)	2.61	1.23	24.7%
MGR USE (n = 738; coefficient $\alpha = 0.93$)			
To what extent do you use performance measurement information for the following activities? ^a			
1. Setting program priorities	3.82	1.03	68.8%
2. Allocating resources	3.75	1.07	66.0%
3. Adopting new program approaches or changing work processes	3.78	1.04	66.9%
4. Coordinating program efforts with other internal or external organizations	3.59	1.08	59.6%
5. Refining program performance measures	3.67	1.12	61.9%
6. Setting new or revising existing performance goals	3.74	1.09	65.6%
7. Setting individual job expectations for government employees I manage or supervise	3.68	1.09	64.5%
8. Rewarding government employees I manage or supervise	3.62	1.12	60.1%
HIGHER USE (n = 624; coefficient $\alpha = 0.87$)			
To what extent do you agree with the following statements? ^a			
1. Results-oriented performance information from my activities is used to develop my agency's budget	2.92	1.15	28.9%
2. Funding decisions for my activities are based on results-oriented performance information	2.78	1.12	23.5%
3. Changes by management above my level are based on results-oriented performance information	2.68	1.14	23.1%
RESULTS TO DATE (n = 501)			
1. To what extent do you believe that your agency's efforts to implement GPRA <i>to date</i> have improved your agency's programs/operations/projects? ^a	2.45	1.03	13.7%
FUTURE RESULTS (n = 596)			
1. To what extent do you believe that implementing GPRA can improve your agency's programs/operations/projects <i>in the future</i> ? ^a	3.08	1.10	34.7%

TABLE 1 (continued)

Construct and Survey Items	Mean	Std. Dev.	%Great or Very Great Extent^b
DATA LIMITATIONS (n = 685; coefficient $\alpha = 0.84$)			
To what extent have the following factors <u>hindered</u> measuring performance or using performance information? ^a			
1. Difficulty obtaining valid or reliable data	3.00	1.23	38.1%
2. Difficulty obtaining data in time to be useful	2.80	1.23	29.6%
3. High cost of collecting data	2.60	1.26	25.0%
4. Existing information technology not capable of providing needed data	2.61	1.26	26.6%
METRIC DIFFICULTIES (n = 701; coefficient $\alpha = 0.81$)			
To what extent have the following factors <u>hindered</u> measuring performance or using performance information? ^a			
1. Difficulty determining meaningful measures	3.36	1.21	48.1%
2. Results of our program(s)/operation(s)/project(s) occurring too far in the future to be measured	2.39	1.24	19.6%
3. Difficulty distinguishing between the results produced by the program and results caused by other factors	2.68	1.17	23.3%
4. Difficulty determining how to use performance information to improve the program	2.48	1.12	18.5%
5. Difficulty determining how to use performance information to set new or revise existing performance goals	2.45	1.13	19.0%
COMMITMENT (n = 611; coefficient $\alpha = 0.65$)			
1. To what extent does your agency's <u>top leadership</u> demonstrate a strong commitment to achieving results? ^a	3.61	1.19	62.88
2. To what extent has the lack of ongoing <u>top executive</u> commitment or support for using performance information to make program/funding decisions <u>hindered</u> measuring performance or using performance information? ^a (reverse coded in the construct)	2.30	1.25	18.92
3. To what extent has the lack of ongoing <u>congressional</u> commitment or support for using performance information to make program/funding decisions <u>hindered</u> measuring performance or using performance information? ^a (reverse coded in the construct)	2.66	1.41	31.77
AUTHORITY (n = 765)			
1. Agency managers at my level have the decision making authority needed to help the agency accomplish its strategic goals ^a	3.07	1.07	37.30
TRAINING (n = 747)			
During the past 3 years, has your agency provided, arranged, or paid for training that would help you to accomplish the following tasks? (1 = yes, 0 = no):			
1. Conduct strategic planning	0.50	0.50	n/a
2. Set program performance goals	0.46	0.50	n/a
3. Develop program performance measures	0.42	0.49	n/a
4. Use program performance information to make decisions	0.38	0.48	n/a
5. Link the performance of program(s)/operation(s)/project(s) to the achievement of agency strategic goals	0.40	0.49	n/a
GPRA INVOLVEMENT (n = 756; coefficient $\alpha = 0.91$)			
To what extent have you and your staff been involved in your agency's efforts in implementing GPRA? ^a			
1. Your involvement.	2.48	1.31	23.5%
2. Your staff's involvement.	2.19	1.28	17.3%

- a. Scale: 1 = no extent, 2 = small extent, 3 = moderate extent, 4 = great extent, 5 = very great extent. Reported sample sizes and coefficient alphas are for observations with responses to all of the questions used to compute the respective constructs.
- b. The percentage of respondents answering "to a great extent" or "to a very great extent".

Table 2

Spearman Correlations Among the Implementation Factor, Measurement System Development, and System Outcome Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. MEASUREMENT	1.00													
2. ACCOUNTABILITY	0.47***	1.00												
3. MGR USE	0.54***	0.40***	1.00											
4. HIGHER USE	0.47***	0.47***	0.39***	1.00										
5. RESULTS TO DATE	0.39***	0.29***	0.39***	0.47***	1.00									
6. FUTURE RESULTS	0.14***	0.09**	0.25***	0.24***	0.60***	1.00								
7. DATA LIMITATIONS	-0.21***	-0.24***	-0.09**	-0.07*	-0.01	0.14***	1.00							
8. METRIC DIFFICULTIES	-0.41***	-0.37***	-0.32***	-0.23***	-0.24***	-0.06	0.57***	1.00						
9. COMMITMENT	0.38***	0.58***	0.29***	0.41***	0.30***	0.11**	-0.28***	-0.44***	1.00					
10. AUTHORITY	0.39***	0.58***	0.33***	0.46***	0.37***	0.17***	-0.10***	-0.22***	0.44***	1.00				
11. TRAINING	0.29***	0.24***	0.23***	0.30***	0.31***	0.14***	0.02	-0.12***	0.24***	0.25***	1.00			
12. PILOT	0.09*	0.01	0.02	0.06*	0.07*	0.01	0.004	-0.01	-0.03	-0.03	0.03	1.00		
13. SES	0.18***	0.14***	0.16***	0.13***	0.06	-0.02	0.01	-0.06*	0.23***	0.21***	0.24***	0.02	1.00	
14. GPRA INVOLVEMENT	0.35***	0.24***	0.27***	0.29***	0.42***	0.22***	0.07*	-0.06	0.25***	0.30***	0.39***	0.14***	0.46***	1.00

MEASUREMENT = the extent to which results-oriented performance measures have been developed and implemented; ACCOUNTABILITY = the extent to which managers are held accountable for achieving results; MGR USE = the use of performance data by managers for decision-making; HIGHER USE = the use of performance information for higher-level agency or funding decisions; RESULTS TO DATE = the perceived extent the U.S. Government Reporting and Results Act (GPRA) has positively influenced agency performance; FUTURE RESULTS = the perceived extent the GPRA will positively influenced agency performance in the future; DATA LIMITATIONS = the extent information system or data problems hinder performance measurement; METRIC DIFFICULTIES = the extent problems identifying, developing, and assessing appropriate performance metrics hinder performance measurement; COMMITMENT = management commitment to performance measurement; AUTHORITY = respondents' decision-making authority; TRAINING = training in performance measurement and use of performance information; PILOT = GPRA pilot site; SES = member of the Senior Executive Service; and GPRA INVOLVEMENT = the extent respondent or staff is involved in implementing the GPRA's requirement.

***, **, * indicate statistical significance at the one, five, and ten percent levels (two-tailed), respectively.

Table 3

Determinants of Results-Oriented Performance Measure Development
by U.S. Government Managers

	Hypothesized	
	Sign	MEASUREMENT
DATA LIMITATIONS	-	-0.02 (-0.36)
METRIC DIFFICULTIES	-	-0.28*** (-5.59)
COMMITMENT	+	0.11** (2.40)
AUTHORITY	+	0.20*** (5.74)
TRAINING	+	0.07*** (4.63)
PILOT	+	0.13** (2.12)
SES	?	0.08 (1.33)
Adjusted R ²		0.30
F		33.04***
Sample Size		528

Ordinary least squares coefficients, with corresponding t-statistics in parentheses. Intercept terms are not reported. MEASUREMENT = the extent to which results-oriented performance measures have been developed and implemented; DATA LIMITATIONS = the extent information system or data problems hinder performance measurement; METRIC DIFFICULTIES = the extent problems identifying, developing, and assessing appropriate performance metrics hinder performance measurement; COMMITMENT = management commitment to performance measurement; AUTHORITY = respondents' decision-making authority; TRAINING = training in performance measurement and use of performance information; PILOT = GPRA pilot site; SES = member of the Senior Executive Service.

***, **, * indicate statistical significance at the 1%, 2.5% and 5% levels, respectively. Significance levels are one-tailed for predictor variables with hypothesized signs and two-tailed for control variables.

Table 4

Determinants of U.S. Government Managers'
Accountability for Achieving Results

	Hypothesized Sign	ACCOUNTABILITY
MEASUREMENT	+	0.15*** (4.58)
DATA LIMITATIONS	-	-0.08** (-2.40)
METRIC DIFFICULTIES	-	-0.03 (-0.90)
COMMITMENT	+	0.29*** (8.46)
AUTHORITY	+	0.27*** (10.10)
TRAINING	+	0.02* (1.76)
PILOT	0	-0.01 (-0.13)
SES	?	-0.05* (-2.15)
Adjusted R ²		0.51
F-statistic		69.81***
Sample Size		524

Ordinary least squares coefficients, with corresponding t-statistics in parentheses. Intercept terms not reported. ACCOUNTABILITY = the extent to which managers are held accountable for achieving results; MEASUREMENT = the extent to which results-oriented performance measures have been developed and implemented; DATA LIMITATIONS = the extent information system or data problems hinder performance measurement; METRIC DIFFICULTIES = the extent problems identifying, developing, and assessing appropriate performance metrics hinder performance measurement; COMMITMENT = management commitment to performance measurement; AUTHORITY = respondents' decision-making authority; TRAINING = training in performance measurement and use of performance information; PILOT = GPRA pilot site; SES = member of the Senior Executive Service.

***, **, * indicate statistical significance at the 1%, 2.5% and 5% levels, respectively. Significance levels are one-tailed for predictor variables with hypothesized signs and two-tailed for control variables.

Table 5

Determinants of the Use of Results-Oriented Performance Information
by U.S. Government Managers

	Hypothesized		
	Sign	MGR USE	HIGHER USE
MEASUREMENT	+	0.45*** (10.44)	0.30*** (6.20)
ACCOUNTABILITY	+	0.15*** (2.69)	0.23*** (3.47)
DATA LIMITATIONS	-	0.14*** (3.26)	0.11** (2.31)
METRIC DIFFICULTIES	-	-0.19*** (-3.83)	-0.004 (-0.08)
COMMITMENT	+	-0.04 (-0.75)	0.15*** (2.66)
AUTHORITY	+	0.03 (0.74)	0.18*** (4.23)
TRAINING	+	0.02 (1.07)	0.05*** (2.91)
PILOT	0	-0.08 (-1.23)	-0.03 (-0.47)
SES	?	0.08 (1.35)	-0.06 (-0.88)
Adjusted R ²		0.37	0.38
F-statistic		34.32***	32.95***
Sample Size		508	472

Ordinary least squares coefficients, with corresponding t-statistics in parentheses. Intercept terms not reported. MGR USE = the use of performance data by managers for decision-making; HIGHER USE = the use of performance information for higher-level agency or funding decisions; MEASUREMENT = the extent to which results-oriented performance measures have been developed and implemented; ACCOUNTABILITY = the extent to which managers are held accountable for achieving results; DATA LIMITATIONS = the extent information system or data problems hinder performance measurement; METRIC DIFFICULTIES = the extent problems identifying, developing, and assessing appropriate performance metrics hinder performance measurement; COMMITMENT = management commitment to performance measurement; AUTHORITY = respondents' decision-making authority; TRAINING = training in performance measurement and use of performance information; PILOT = GPRA pilot site; SES = member of the Senior Executive Service.

***, **, * indicate statistical significance at the 1%, 2.5% and 5% levels, respectively. Significance levels are one-tailed for predictor variables with hypothesized signs and two-tailed for control variables.

Table 6
Determinants of the Perceived Benefits From the
U.S. Government Performance and Results Act (GPRA)

	Hypothesized		
	Sign	RESULTS TO DATE	FUTURE RESULTS
MEASUREMENT	+	0.26*** (3.36)	0.01 (0.11)
ACCOUNTABILITY	+	-0.07 (-0.70)	-0.17* (-1.68)
DATA LIMITATIONS	-	0.21*** (3.07)	0.32*** (4.36)
METRIC DIFFICULTIES	-	-0.22*** (-2.74)	-0.19** (-2.21)
COMMITMENT	+	0.13 (1.58)	0.03 (0.30)
AUTHORITY	+	0.12** (1.95)	0.16*** (2.52)
TRAINING	+	0.07*** (3.02)	0.06** (2.17)
PILOT	0	0.09 (0.85)	-0.11 (-1.01)
GPRA INVOLVEMENT	?	0.28*** (4.53)	0.25*** (3.84)
SES	?	-0.17 (-1.62)	-0.28*** (-2.55)
Adjusted R ²		0.27	0.12
F-statistic		15.30***	6.86***
Sample Size		380	434

Ordinary least squares coefficients, with corresponding t-statistics in parentheses. Intercepts not reported. RESULTS_TO_DATE = the perceived extent the U.S. Government Reporting and Results Act (GPRA) has positively influenced agency performance; FUTURE RESULTS = the perceived extent the GPRA will positively influenced agency performance in the future; MEASUREMENT = the extent to which results-oriented performance measures have been developed and implemented; ACCOUNTABILITY = the extent to which managers are held accountable for achieving results; MGR USE = the use of performance data by managers for decision-making; HIGHER USE = the use of performance information for higher-level agency or funding decisions; DATA LIMITATIONS = the extent information system or data problems hinder performance measurement; METRIC DIFFICULTIES = the extent problems identifying, developing, and assessing appropriate performance metrics hinder performance measurement; COMMITMENT = management commitment to performance measurement; AUTHORITY = respondents' decision-making authority; TRAINING = training in performance measurement and use of performance information; PILOT = GPRA pilot site; SES = member of the Senior Executive Service; and GPRA INVOLVEMENT = the extent respondent or staff is involved in implementing the GPRA's requirement.

***, **, * indicate statistical significance at the 1%, 2.5% and 5% levels, respectively. Significance levels are one-tailed for predictor variables with hypothesized signs and two-tailed for control variables.

Figure 1

Hypothesized Conceptual Model Linking Implementation Factors, Measurement System Development, and System Outcomes

