Virginia Polytechnic Institute and State University

From the SelectedWorks of Matthew Dull

April, 2009

Results-Model Reform Leadership: Questions of Credible Commitment

Matt Dull, Virginia Tech

Available at: https://works.bepress.com/matthew_dull/2/
Results-Model Reform Leadership: Questions of Credible Commitment

Matthew Dull
Center for Public Administration and Policy—Virginia Tech

ABSTRACT

Few problems common in management reform are more prominent or more vaguely conceived than is leadership. Advocates and observers broadly agree that leadership is a critical factor where reform takes hold. Yet, in scholarship assessing the results-model management reforms proliferating in public organizations during the last decade and a half, leadership remains an elusive concept, rarely subject to empirical scrutiny. Applying the logic of credible commitment drawn from the study of institutional political economy, this article models leadership commitment as a factor shaping organizational responses to reform. Quantitative analysis of data drawn from two Government Accountability Office surveys of agency managers administered during the implementation of the Government Performance and Results Act provides evidence regarding the impact of leadership commitment on perceived credibility results-based reform and reported use of performance measures. The article concludes with a brief discussion of reputation-based credibility and the skepticism many government managers hold toward reform.

INTRODUCTION

The broad diffusion of results-model reform initiatives, centering on the use of performance measures to improve governance, has in recent years given rise to an empirical research literature examining their adoption and implementation. It will surprise few observers that many of the most intriguing research questions emerge not from performance management’s lauded successes but in its many disappointments. Why have results-model reforms so often failed to take hold? No doubt there as many explanations for failure as there are failed initiatives. Yet, one factor—leadership commitment—surfaces so often in conversation and analysis that something approaching a professional consensus holds it
essential for organizational change. Summarizing the Government Performance Project and Federal Performance Project assessments of results-model reforms across city, state, and federal agencies Ingraham, Joyce, and Donahue note, “. . .we had not initially intended to study leadership—indeed, were advised not to do so.” They explain:

Instead, we focused on the ability of governments and agencies to manage toward results, to build effective management systems, and to align and integrate them in productive ways. Yet in every case we found strength in these areas, we also found a strong leader or leadership team. Conversely, those governments and agencies that struggled with management and performance were notable for the absence of leadership and direction. (Ingraham, Joyce, and Donahue 2003, 131, 152)

As a matter of social science, leadership has a reputation for being vague and idiosyncratic—a good subject to avoid. Nonetheless, leadership looms large in the minds of managers compelled to invest scarce time and attention implementing reform. Leadership and managerial commitments bear a well-documented influence in the adoption of results-model reform. Where research shifts to examining the impact of these reforms, leadership commitment is still more prominent. This article contributes to developing scholarship on leadership and reform with an analytical treatment and quantitative analysis of available empirical evidence.

The model developed here centers on the relationship between public sector leaders promoting reform and the middle- and upper-level managers whose actions form a vital link between reform and organizational performance. Motivating the analysis is a dilemma familiar to students of performance-based management: Despite decades of research mapping the causes and consequences of “misaligned” or nonproductive responses to performance incentives, these behaviors remain unmistakable features of
performance management. The economic rationale of performance measurement—rooted in models of principals and agents under incomplete contracting—has given rise to a complex architecture of formal theories examining the precision of incentives and the mechanics of performance monitoring (Alchian and Demsetz 1972; Brickley, Smith, and Zimmerman 2001). In Managerial Dilemmas, Miller (1992) demonstrates the unfinished nature of this economic rationale. Whereas principal-agent theory has traditionally focused on the agent’s capacity for moral hazard, Miller’s “vertical dilemma” suggests moral hazard is a two-way street. The problem of the performance contract is not only a matter of aligned incentives, but of aligned incentives constructed on the solid foundation of credible, purposeful executive authority (Miller 2000, 297). In this sense, public sector results-based management reform and the broader agenda advanced under the banner of New Public Management rely on an often unspoken appeal to what in the Barnard-Simon tradition of organization theory is known as the zone of acceptance. Within the zone of acceptance, individuals accede to credible executive discretion. Its boundaries define a tension so firmly embedded in contemporary organizational thinking that it is often taken for granted. The tension between individual choice and collective action governed by hierarchical authority is precisely what many contemporary advocates of performance measurement neglect—and precisely where one often finds the source of reform’s undoing.

Miller (1992) models this classic dilemma as a problem of credible commitments in the use of political authority, a concept drawn from contemporary institutional analysis. The problem of credible commitment offers a useful heuristic by which to examine the relationship between leadership and reform. How credible are leaders in their reform commitments? What are the terms on which credibility is defined? Managers and employees form discernable judgments about leaders’ credibility and rely on those judgments to guide their own behavior. Prior experience often leaves managers feeling reform initiatives waste time and effort, with little consequence for organizational performance. At the center of contemporary reform are questions about the credibility of the commitments policy makers and administrators make to change how government agencies operate. Stated succinctly, credibility is a necessary condition for performance management. If managers fear that leadership is not credible—if they suspect leaders lack competence, authority, or trust—the capacity to produce credible information will not take hold.

Following the credible commitment logic, this article presents empirical evidence highlighting the influence of perceived leadership commitment on the credibility of results-model reform and the reported use of performance measures in public management. Quantitative data are drawn from surveys of middle- and upper-level (GS-13 to

---

4 Thus, phrases like “cream skimming” or “teaching to the test” have entered the public management vernacular. Gormley and Weimer (1999) provide a review of what they term “response problems” applied to organizational report cards. Kerr (1975) remains a classic statement of misaligned incentives. Galbraith and Merrill (1996) offer an empirical review, characterizing organizational environments as nonmanipulative, selectively manipulative, and highly manipulative. Among the most robust evidence in the US public sector comes from the administration of federal job training programs (Heinrich 2002; Marschke 2001).

5 Miller builds this proposition from an impossibility theorem developed by economist Holmstrom (1982). Holmstrom’s proof demonstrates that given (1) an interdependent production process, meaning the marginal productivity of every group member depends on the effort of each of the other group members and (2) information asymmetry, meaning only the overall output is observable by outside monitors—it is impossible to develop an allocation that is stable, budget balancing, and Pareto Optimal (Miller 1992, 129, 2000, 294). Miller (2005) provides a useful review of this and other theoretical developments shaping contemporary principal-agent models.

6 This is Simon’s (1947) terminology. Barnard (1938) employs the less poetic “zone of indifference.”
GS-15/Senior Executive Service [SES]) US Federal government managers administered by Government Accountability Office (GAO) in 1996–97 ($N = 905$) and 2000 ($N = 2,150$), at two key points in agency implementation of the Government Performance and Results Act (GPRA). The credible commitment framework and the empirical findings presented here point to potentially fruitful extensions where theory and observation are guiding the integration of principles drawn from political economy and empirical organizational research. Moreover, the logic of credible commitment points to the relevance of scholarship to practice and the agency politics managers and administrators observe every day.

**LEADERSHIP, COMMITMENT, AND REFORM**

Consider for a moment why some economies succeed while others fail. In recent decades, this question has given rise to a body of research organized around the following proposition: Economic prosperity depends in part on the degree to which actors are confident that existing economic arrangements—“the rules of the game”—will not be altered arbitrarily in the future. Where property rights are inadequately enforced or subject to the whim of political authorities, market activity and economic growth are inhibited (North 1981; Torstensson 1994). The story of a sovereign and her constituents illustrates the principle. Imagine the constituents possess a small endowment of seeds for planting a harvest. At any point in time, the sovereign faces a hypothetical choice between maximizing short-run gains by confiscating whatever economic resources her constituents possess or seeking to maximize her gains over time by allowing her constituents to plant seeds, grow crops, and taxing what they produce. Imagine she favors the latter and a rate of tax is agreed. A season passes, the constituents raise their crops and the harvest arrives. Again, the sovereign faces a choice: Collect the agreed tax or maximize short-run gains by seizing more of her constituents’ economic resources than the agreed tax permits. This unrivaled power presents the sovereign with a dilemma. Her constituents are aware of this trade-off. Fearing the sovereign will act to maximize short-term gains, they may respond by pursuing short-run strategies that protect their own resources but reduce overall wealth (Miller 1992, 154–7; North 1994). As a result, the sovereign may benefit by committing to a set of governing arrangements—an established, credible system of property rights—that will foster overall prosperity.

The logic of credible commitment holds that political authorities may benefit by restricting their own future actions. Their authority may be bound to existing governance by formal constraints or “tying one’s hands,” or they may establish a reputation such that shared expectations hold their commitments credible. Both strategies purchase credibility, reducing strategic uncertainties but at a cost. Because they are comparatively reliable, formal constraints are potent instruments for establishing credibility. Yet, by restricting future choices, formal constraints may carry unforeseen costs, reduce adaptive capacity,  

7 In a seminal treatment, North and Weingast (1989) observe a variation on this dilemma in the 17th English monarchy. Pressed by the financial burdens of war making, the monarchy came to rely on the frequent infusion of economic resources through market interventions, in the process creating an economic climate that discouraged savings and investment. Revolt and the rise of Parliament did little to diminish the threat of political violations of property rights and the English economy continued to suffer. With the Glorious Revolution of 1688, these economic costs were recognized and new constraints on political interventions in the economy were established, laying a foundation of credible property rights that served as the basis for England’s sustained economic expansion.
or prove unenforceable. Reputational commitments, on the other hand, require time to cultivate and may be discounted as “cheap talk.”⁸ Political and organizational circumstances may limit the capacity of leaders to develop and benefit from reputation-based credibility. A short or uncertain time horizon enhances the threat of cheap talk and, as a consequence, the relative value of formal constraints.

In the US federal agencies, turnover among political administrators, competition among principals, and shifting political winds often contribute to short time horizons, sharpening skepticism about reform promises. At the same time, consider how commonly public sector reform leaders depend on a cultivated reputation for “responsible behavior” (Miller 1992, 220; North and Weingast 1989).⁹ This is not only because of the demonstrated power of normative commitments even where the substance of reform is opposed (Golden 1999) but because formal and informal authority are only analytically separable.¹⁰ Credibility is a quality not of an individual or action but of an individual action in relationship to others. Whether a constitutionally bound sovereign or a professional civil servant bound to normative commitments, credibility remains fundamentally relational. In the traditional of organization research following from the work of Barnard (1938) and Simon (1947), this relational character defines a zone of acceptance within which individuals accede to executive discretion. The boundaries of zone of acceptance are shaped by the credibility of actions and communications; indeed, for Barnard, neither executive authority nor cooperation around common purposes logically precedes the other.¹¹ Contemporary treatments like Miller’s credible commitment or Kreps’s (1990) treatment of corporate culture begin from the propositions that organizations represent repeat interactions, formal strategies yield multiple equilibria, and formal contracts are expensive to establish and monitor. In the political economy of organizations following from these propositions, the

---


⁹ Think of James Lee Witt’s first day as the director of the Federal Emergency Management Agency (FEMA), standing in the front lobby greeting FEMA’s politics-weary employees as they arrived at work; or Witt’s open telephone line on Tuesday mornings demonstrating his commitment to the organization’s success (Khademian 2002, 51–2). Consider, at least by reputation, Witt’s capacity to fuse that credibility into an executive authority that helped to foster FEMA’s organizational renewal. This model of professional, reputation-based commitment resembles what in the field of organizational behavior is termed the psychological contract (Guest 2004; Rousseau and Tijoriwala 1998). Barbara Romzek invokes the psychological contract in her account of public sector employee commitment rooted in the credibility of the employer’s long-term reputation. Commitment is a, “two-way street,” Romzek observes, “agencies and supervisors must demonstrate commitment to get commitment” (Romzek 1992, 153, 149).

¹⁰ The blunt distinction between formal and informal commitments overlays what are, in fact, rich variations in flexibility and concreteness across different forms of social organization. Arthur Stinchcombe (2001) persuasively develops this point, arguing further that flexibility in a broad array of social arrangements depends critically on complementary rigid forms. The legislative establishment of the merit system, for example, defined a set of formal, legal commitments designed to enhance bureaucratic competence by buffering the civil service from the push and shove of politics (Miller 2000, 316). However, an account of political buffering and bureaucratic control in the federal agencies that focuses exclusively on formal administrative architecture is incomplete (Brehm and Gates 1997; Feldman 1989).

¹¹ Barnard (1938, 86) writes, “A purpose does not incite cooperative activity unless it is accepted by those whose efforts will constitute the organization. Hence, there is initially something like simultaneity in the acceptance of a purpose and willingness to cooperate.” James Kouzes and Barry Posner offer a popular contemporary treatment emphasizing the link between credibility and leadership. Based on a survey Kouzes and Posner have administered to individuals across settings, they conclude, “What we found in our investigation of admired leadership qualities is that more than anything, people want leaders who are credible. Credibility is the foundation of leadership.” (Kouzes and Posner 2002, 32–3; original italics).
social character of reputation and credibility are central dynamics in organizational governance.12

Results-model reforms like GPRA embody a mix of formal and informal commitments. The strategic planning prescribed under GPRA aims to secure a sense of common purpose and credible authority. GPRA’s measurement mandates are modeled on the logic of performance contracts, but embedded within this formal architecture are the reform ideals Light terms “liberation management” (Light 1997, 233–5). Together, they propose a performance bargain: greater accountability for results but new flexibilities in achieving them. Is it a credible bargain? In practice, results-model management reforms often have generated only half of the performance bargain. New technologies of control have been created, new data reported, but without enhanced flexibility (Moynihan 2005, 2006).13 Just as often, these reforms are ultimately perceived as nothing more than paperwork exercises, adding new layers of administrative sediment (Light 1997).

The practical knowledge that enables managers to accomplish routine work and prioritize multiple often ambiguous or ill-structured demands is populated by shared categories and simple rules, illustrated in the area of reform by the common suspicion that reform is nothing more than the “flavor of the month.” Because it is so useful, this common wisdom represents among the most formidable tests for reform initiatives. Manager resistance to performance measurement commonly centers on a few frequently repeated “automatic” objections.14 At the organization level, this is often indistinguishable from the resistant cultures and competencies that regularly slow change and frustrate reform ambitions (Hill and Hupe 2002; Kelman 2005; Levitt and March 1995).15 If they perceive a lack of credible leadership commitment to results-model reform, managers may see reform as a threat or a nonissue, gathering less information from fewer, less diverse sources, engaging fewer people, and leaning on familiar ideas and practices.16 Dedicated the cognitive resources necessary to alter routine performance represents a risk for managers for the simple reason that reform promises may not materialize. To highlight the empirical evidence supporting the credible commitment logic, the following sections

12 Similarly, the common economic treatment of reputation as the “record of one’s past performance” invites the question: according to whom? Reputations exist not as qualities of individuals but perceptions of those individuals represented in the minds of people around them. They are socially and politically constituted (Fombrun and Shanley 1990; Podolny 1993; March and Heath 1994). Klein (1997) provides a useful overview of reputation as an economic concept.

13 Wilson (1989, 228) observes the tendency in military history of commanders to employ new technologies—from the telegraph to the computer—to reduce the discretion of frontline officers.


15 Consider a recent study of outcomes measurement in federal agencies complicated because many managers could not distinguish outcomes measures, even when they collected and reported on them (Hatry et al. 2003). It is a striking observation, given the steady drumbeat echoing through government agencies in recent years. One is tempted to echo Schick’s skeptical appraisal of the broader literature, “Surely, there is something amiss when, after decades of patient explanation, just about every treatment of this subject has to present a set of basic definitions, as if words such as inputs and outputs are so obtuse as to defy ordinary understanding” (Schick 2001, 40).

16 Managers believing the organization and its leaders are committed to results-model reform may be more likely to seek new information, open decision processes, revise or amend measures as necessary, and integrate the principles of the results-model in their actions and thinking (Boynton, Gales, and Blackburn 1993; Jackson and Dutton 1988). Behn (2002) develops a parallel treatment.
develop a quantitative analysis evaluating the impact of leadership commitment on manager perceptions of and attention to results-model reform under GPRA.

**QUANTITATIVE ANALYSIS**

The link between leadership commitment and the credibility of reform is examined here using data drawn from two GAO-administered surveys of US federal agency managers GS-13 to GS-15 and SES on performance measurement under GPRA. Though they represent reported rather than observed conditions, these data yield indicators of the perceived credibility of GPRA’s reform mandates and the reported use of performance measures, as well as controls for a variety of organizational, task-contingent, and environmental factors. Where the success and failure are attributed to leadership, perhaps, we may find a more straightforward explanation in agency context or program type. Certain types of government programs—such as international aid programs or programs focused on research and development—face distinct technical and political obstacles to outcomes measurement (Frederickson and Frederickson 2006). Administrators may find their organization lacks the information systems or expertise required to collect and analyze information gauging organizational outcomes. The program’s mission and goals may be characterized by ambiguity or conflicting values, thus complicating and increasing the risks associated with outcomes measurement. Faced with these challenges, administrators may concede that measurement mandates like those under GPRA will generate only negative attention. They may view GPRA as likely to fail or simply look for more promising opportunities to dedicate their limited time and attention. The quantitative models developed here examine the influence of leadership commitment in managers’ perceptions of results-model reforms while controlling for several confounding factors commonly associated with the usefulness of results-model management. Descriptive statistics and information about the construction of the variables are included in the article’s appendix.

**DEPENDENT VARIABLES**

Three dependent variables specified in the models—Credible, Use, and Useful—are summarized in figure 1. Credible gauges respondents’ belief that GPRA will improve how their agencies operate. This variable sums two adjacent survey items, gauging: first, if the implementation of GPRA had “to date” improved their agency’s performance and second if implementation of GPRA “can” improve agency performance. These subjective assessments are combined in the Credible variable, to indicate respondents’ disposition toward reform. To what degree do they believe the claims of results-model reform advocates are credible? The second dependent variable, labeled Use, is an additive index drawn from a six-item series gauging reported use of performance information across a variety of management functions. Does the respondent use the information obtained from performance measurement when setting program priorities, allocating resources, changing work processes, refining performance measures, setting expectations, and rewarding employees? A third dependent variable, labeled Useful, sums five items included in the 2000 survey, gauging the degree to which types of performance measures—output, efficiency, service, quality, and outcome measures—are perceived useful for managing the respondents’ programs or projects in which the respondents are involved. The
Cronbach’s alpha values presented in figure 1 suggest all three dependent variables index well.

Figure 1 highlights a few notable features of the variable distributions. First, between 1997 and 2000, the *Use* variable, measuring reported usage of performance measures, decreases. When GAO (2002) reported the 2000 survey results, this finding, in particular, drew broad attention and concern among agency reform advocates. There is some reason to be skeptical that an objective decrease in the use of performance measures occurred between the two surveys. Rather the differences may be a consequence of differing samples or, perhaps, the attention to performance measurement that accompanied the implementation of GPRA altered the expectations by which respondents evaluated the use of performance measures. The two samples are analyzed separately here, focusing on the relationship between reported use and a handful of variables gauging perceptions of leadership and context at two points in time. A second, startling feature of the distributions is the spike—9.5% of 1997 respondents—registering the maximum total score on the *Use* variable. Almost one in 10 respondents reported the highest possible levels on all six of the items that make up the *Use* index. This proportion, which falls to 3.6% in 2000, suggests one of the critical challenges in the use of subjective reporting, particularly when the survey’s audience is GAO, an influential monitor of agency management. Are these very optimistic respondents engaged in strategic reporting, perhaps, in order to influence external perceptions of the agency’s performance under GPRA? The data analyzed here can neither decisively confirm nor reject this suspicion. However, as discussed below, models omitting likely “strategic” responses yield substantively similar estimates. The findings reported here are robust across a range of alternative specifications of the dependent variables, including models predicting the individual survey items comprising the indices.

Particularly noteworthy in figure 1 is the proportion of values underlying the *Credible* variable and to a lesser extent the *Use* and *Useful* variables falling in the “no basis” (NB)
category. GAO’s survey instrument follows the methodological norm of providing a “no basis to judge” response option for each of the survey’s substantive questions. This practice follows from a long-standing concern among survey researchers that absent such an option respondents may feel pressured to report an opinion. Where respondents have not formed a prior judgment, they may report so-called nonattitudes, thus potentially compromising the analysis (Converse 1964). The prevalence of NB responses given the challenging cognitive task of evaluating the usefulness of performance measurement is no surprise. However, the large proportion of missing values that result creates a secondary problem. Applying “list-wise deletion,” missingness in the Credible variable alone reduces both data sets by almost 40%, dropping the 1997 data set from roughly \( N = 900 \) to \( N = 550 \) and the 2000 data set from approximately \( N = 2,500 \) to \( N = 1,450 \). If the cases dropped from the analysis are not missing at random—that is, if the likelihood of an NB response is correlated with one or more of the explanatory variables—the estimates of the coefficients may be biased (Allison 2002). In the accompanying statistical analysis, a Heckman selection model is employed to evaluate and control for correlation between the explanatory variables and the likelihood respondents answered NB to the questions that form the dependent variables.

**INDEPENDENT VARIABLES**

The causal diagram in figure 2 highlights the independent variables included in the analysis and the basic structure of the Heckman model. The first explanatory variable, Leadership Commitment, draws on a single survey item gauging the perceived commitment of agency leaders to “achieving results.” The basic thesis outlined above holds that Leadership Commitment should yield positive estimates predicting each of the dependent variables. All else equal, high levels of perceived Leadership Commitment to results-model reform will be associated with both higher levels of reported use of performance measures and greater confidence that results-model reform will improve organizational performance.

Two variables gauge perceptions regarding the influence of organizational context and practices on results-model reform. Management combines two survey items indicating the degree to which existing management practices reflect results-model reform principles: holding managers accountable for results and recognizing employees for the accomplishment of strategic goals. Information combines survey items gauging the degree to which managers believe deficiencies in their organization’s information systems or existing staff expertise will hinder performance measurement. The scale of the Information variable is reversed, so that more positive assessments regarding the agency’s information capacity and expertise are associated with higher values. As figure 2 indicates, both the Management and Information variables are expected to generate positive coefficients in models. Due to similarities between items underlying the Management independent variable and the Use and Useful dependent variables, the Management variable is omitted from models predicting these variables.

The Measurement Challenges variable gauges perceived technical obstacles to developing and using performance measures. It is drawn from the most pronounced of several factors believed to hinder the development of measures that are valid, reliable, timely, and useful. It is expected to generate negative coefficients. Also included in the 1997 models, though unavailable in 2000, are a series of dichotomous variables controlling for Program Type. “Service delivery” programs are employed as the omitted category. A separate
A dichotomous variable gauging Defense/Nondefense agency is included in the 1997 models as well. Two variables gauge the influence of perceived agency or program political environment. The first, Political Conflict, registers the degree to which respondents believe conflict in Congress and among internal and external agency stakeholders hinders performance measurement within their agencies. A second variable, Congressional Neglect, gauges whether respondents believe lack of attention in Congress hinders the development of performance measures. The Political Conflict and Congressional Neglect variables are expected to generate negative coefficients.

Knowledge of GPRA is based on a single survey item reporting involvement and attentiveness to performance measurement and management reform. Based on the presumption that prior awareness of GPRA reflects attentiveness or involvement in results-based reform, this variable is expected to generate positive coefficients. Several additional variables control for respondent rank and responsibility. A variable gauging the Number of Employees a manager supervises based on a five-point ordinal scale is included. Another five-point scale gauging the respondent’s Grade Level is included in the 1997 models, but is unavailable in 2000. Instead, a five-point variable gauging respondents’ Years as a Supervisor, unavailable in the 1997 survey data, is included in the 2000 models. Though

---

17 Due to changes in the survey instrument, the questions underlying the Political Conflict variable are somewhat different in the 1997 and 2000 models. Differences are described briefly in the article’s appendix.
these measures of supervisory authority, tenure, and rank control key aspects of respondents’ perspectives, no firm hypotheses regarding the influence of these factors were developed. Summary statistics for each of the explanatory variables are presented in table 3 in the article’s appendix.

ESTIMATION: HECKMAN SELECTION MODEL

A maximum likelihood Heckman selection model is employed to control and evaluate the influence of the explanatory variables on missingness due to NB responses in the dependent variables. The Heckman technique generates two sets of coefficients. The first estimates a probit “selection” model with a dichotomous dependent variable equaling 1 if the respondent answered the questions and 0 otherwise.18 The residuals produced by the first-stage estimates generate a new variable, the Inverse Mill’s Ratio or Lambda, which is included as a control variable in the second outcome equation (Heckman 1979). The dependent variables are based on positive, ordinal scales but include large enough ranges that the estimates derived from the linear outcome model do not diverge significantly from more specialized models designed to account for violations of ordinary least squares assumptions.19

In addition to correcting selection bias, the Heckman model yields potentially useful information. Research in the field of survey methodology frames the decision to register NB or other nonresponse variants such as “don’t know” or “no opinion” as a function of three factors: cognitive availability or whether a clear answer can be easily retrieved, a judgment about the adequacy of an answer given expectations, and communicative intent or motivation (Beatty and Herrmann 2002). NB respondents may feel uncertain or believe they lack necessary information, and in this sense, the NB category enhances the validity of the substantive measure. However, an NB response may instead indicate ambivalence. The respondent may feel less uncertain than conflicted about the prospects and usefulness of reform. NB respondents may wish to avoid sending an undesirable or unflattering signal or they may satisfice, responding NB to avoid the effort of synthesizing opinions for which they have all the necessary ingredients (Krosnick 2002; Krosnick et al. 2002). Parsing explanations for the large number of NB responses observed in the Credible and to a lesser extent the Use and Useful dependent variables is not feasible here. Reasoning loosely from the leadership commitment framework, however, two sets of expectations can be derived.

18 To identify the Heckman model, a handful of selection variables are included predicting the likelihood a manager responded: reported involvement in developing measures, gathering and analyzing data, using performance goals to assess program success, and in the 2000 models, gauging whether the respondent is located at their agency’s headquarters.

19 The Leadership, Management, and Information variables might be expected to generate organization-level effects and may themselves be correlated with other explanatory variables. A case can be made that each of the organization-level variables should be treated as endogenous. Each of these factors arguably is caused, in part, by political conflict, neglect, and technical measurement challenges. Leaders who observe intense political conflict, little interest at all among political principals, or see significant barriers to measurement are presumably less likely to embrace performance measurement. Organizational capacity, including technical resources and management structure, may also be treated as a function of politics, for example. If so, the least-squares assumptions of unidirectional causal relationships and uncorrelated error terms would be violated. Alternative specifications modeling endogeneity and organization-level effects yield similar results (Asher 1976; Gujarati 1995, chap. 19). Robust standard errors are employed in the accompanying models.
Again, the credible commitment logic holds reform commitments are credible when they reduce uncertainty and when those affected believe the new rules of the game will endure. Evidence leadership commitment is absent or lacks credibility likewise might therefore be expected to reduce uncertainty, signaling reform’s prospects are dim. Thus, if NB responses indicate uncertainty, Leadership Commitment should be a significant predictor, but because the relationship between the variables is expected to take a curvilinear form no directional hypothesis is specified. If, by contrast, NB responses are a function in part of low aspiration or attentiveness, the relationship between perceived leadership and the likelihood of an NB response should be negative. Based on this, latter way of thinking, NB responses in the Credible, Use, and Usefulness variables might indicate a disposition toward reform. A manager reluctant to commit the cognitive resources necessary to register an opinion regarding performance measurement—or who is informed, but unconvinced by the evidence she has been exposed to—is unlikely to commit her own or her subordinates’ time and attention to the implementation of reform. Again, the Heckman model is applied here first to correct for bias introduced by NB responses, but the model also yields information that may be useful in assessing the credible commitment framework. Two expectations are outlined above. The latter predicts Leadership Commitment will produce negative coefficients in the selection equation, whereas the former suggests both positive and negative perceptions will reduce uncertainty and thus increase the likelihood of a response.\textsuperscript{20} The following section first analyzes results from the outcome equations predicting the Credible, Use, and Usefulness dependent variables, then turns to results from the selection models.

RESULTS AND DISCUSSION

Results for models predicting the Credible dependent variable are summarized in table 1 and the Use and Usefulness variables are summarized in table 2. Each of the models report two sets of coefficients: a first column estimating the selection stage of the Heckman model, or the likelihood of a substantive as opposed to a NB response, and a second column of coefficients from the outcome equation predicting the dependent variable. Focusing first on estimates in the outcome equations across tables 2 and 3, the Leadership Commitment variable is positive and robust across specifications, lending support to the credible commitment framework. As figure 3a illustrates, holding other variables at their means and shifting Leadership Commitment from a moderate score of 3 to a maximum score of 5 boosts predicted Use (1997) from the 31st to the 55th percentile and predicted value of Use (2000) from approximately the 25th percentile to just short of the 50th percentile among all respondents.\textsuperscript{21} The minimum Leadership Commitment value 1, on the other hand, yields

\textsuperscript{20} To the degree that existing organizational factors measured by the Management and Information variables serve to reduce respondent uncertainty and compel managers to attend to performance measurement, these variables should also generate positive coefficients in the “selection” equations. Conversely, the three variables expected to undermine the credibility of reform—Measurement Challenges, Political Conflict, and Congressional Neglect—are expected to increase respondent uncertainty and decrease attentiveness, thus yielding negative coefficients in the selection models. Among the variables measuring respondent characteristics, Awareness of GPRA presumably will be associated with higher levels of attentiveness; as will measures of tenure and status—Number of Employees, Grade Level, and Years as Supervisor, which are expected to positively predict the likelihood an opinion was offered in the dependent variables.

\textsuperscript{21} Predicted probabilities and estimated values were calculated using Clarify implemented in Stata 7 by King, Tomz, and Wittenberg’s (2000).
Table 1
Credible Results-Model Reform

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership Commitment</td>
<td>-0.173 (0.083)**</td>
<td>0.465 (0.116)**</td>
<td>-0.049 (0.039)</td>
<td>0.425 (0.067)**</td>
</tr>
<tr>
<td>Management Accountability and Recognition</td>
<td>0.019 (0.108)</td>
<td>0.177 (0.128)</td>
<td>-0.072 (0.056)</td>
<td>0.221 (0.074)**</td>
</tr>
<tr>
<td>Information Capacity and Expertise</td>
<td>0.022 (0.066)</td>
<td>0.228 (0.095)**</td>
<td>0.051 (0.039)</td>
<td>-0.144 (0.052)**</td>
</tr>
<tr>
<td>Measurement Challenges (Linking Inputs to Outcomes)</td>
<td>-0.002 (0.079)</td>
<td>-0.074 (0.107)</td>
<td>0.128 (0.037)**</td>
<td>-0.295 (0.060)**</td>
</tr>
<tr>
<td>Political Conflict</td>
<td>0.134 (0.078)*</td>
<td>-0.153 (0.114)**</td>
<td>-0.020 (0.034)</td>
<td>0.010 (0.052)</td>
</tr>
<tr>
<td>Congressional Neglect</td>
<td>0.005 (0.055)</td>
<td>-0.037 (0.078)</td>
<td>0.015 (0.030)</td>
<td>0.005 (0.038)</td>
</tr>
<tr>
<td>Knowledge of GPRA</td>
<td>0.776 (0.083)**</td>
<td>0.198 (0.252)</td>
<td>0.784 (0.042)**</td>
<td>-0.039 (0.130)</td>
</tr>
<tr>
<td>Number of Employees</td>
<td>-0.051 (0.059)</td>
<td>0.037 (0.068)</td>
<td>-0.012 (0.040)</td>
<td>0.016 (0.046)</td>
</tr>
<tr>
<td>Years as Supervisor</td>
<td>—</td>
<td>—</td>
<td>0.071 (0.033)*</td>
<td>-0.185 (0.052)**</td>
</tr>
<tr>
<td>Grade Level</td>
<td>0.147 (0.079)**</td>
<td>-0.197 (0.116)*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Involved in GPRA—Developing Measures</td>
<td>0.309 (0.192)</td>
<td>—</td>
<td>0.433 (0.109)**</td>
<td>—</td>
</tr>
<tr>
<td>Involved in GPRA—Gathering/Analyzing Data</td>
<td>0.004 (0.186)</td>
<td>—</td>
<td>0.111 (0.099)</td>
<td>—</td>
</tr>
<tr>
<td>Involved in GPRA—Using Measures</td>
<td>0.472 (0.187)**</td>
<td>—</td>
<td>0.411 (0.094)**</td>
<td>—</td>
</tr>
<tr>
<td>Headquarters</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.075 (0.076)</td>
</tr>
<tr>
<td>Defense/Nondefense</td>
<td>-0.002 (0.217)</td>
<td>0.352 (0.333)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Government Support</td>
<td>-0.347 (0.299)</td>
<td>0.114 (0.431)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Agency Support</td>
<td>-0.022 (0.231)</td>
<td>0.198 (0.272)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Research and Development</td>
<td>-0.038 (0.264)</td>
<td>-0.308 (0.323)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Credit</td>
<td>-0.673 (0.378)*</td>
<td>0.035 (0.633)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Grants</td>
<td>0.270 (0.729)</td>
<td>0.119 (0.557)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Regulation</td>
<td>-0.131 (0.232)</td>
<td>0.058 (0.281)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Program—Other</td>
<td>-0.414 (0.369)</td>
<td>1.37 (0.582)**</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>-0.701 (0.591)</td>
<td>—</td>
<td>-0.918 (0.261)</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.587 (0.564)**</td>
<td>3.392 (0.596)**</td>
<td>-2.865 (0.312)**</td>
<td>5.511 (0.794)**</td>
</tr>
<tr>
<td>N</td>
<td>—</td>
<td>572</td>
<td>—</td>
<td>1,859</td>
</tr>
<tr>
<td>p</td>
<td>-0.393 (0.314)</td>
<td>0.510 (0.134)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Chi-square</td>
<td>88.61**</td>
<td>179.06**</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-984.508</td>
<td>-3036.315</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: Dependent variable: Credible. Estimates from maximum-likelihood Heckman regression model with selection. Robust standard errors reported in parentheses. Base category for Program variables is service delivery.

*p < .10, **p < .05.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leadership Commitment</strong></td>
<td>-0.321 (0.049)**</td>
<td>1.486 (0.195)**</td>
<td>0.139 (0.043)**</td>
<td>1.256 (0.145)**</td>
<td>0.160 (0.042)**</td>
<td>1.124 (0.122)**</td>
</tr>
<tr>
<td><strong>Information Capacity and Expertise</strong></td>
<td>-0.077 (0.047)</td>
<td>0.196 (0.194)</td>
<td>0.082 (0.043)*</td>
<td>-0.167 (0.137)</td>
<td>0.077 (0.042)*</td>
<td>-0.061 (0.112)</td>
</tr>
<tr>
<td><strong>Measurement Challenges (Linking Inputs to Outcomes)</strong></td>
<td>-0.105 (0.053)**</td>
<td>-0.553 (0.214)**</td>
<td>-0.010 (0.047)</td>
<td>-1.089 (0.146)**</td>
<td>-0.035 (0.046)</td>
<td>-1.235 (0.126)**</td>
</tr>
<tr>
<td><strong>Political Conflict</strong></td>
<td>0.061 (0.054)</td>
<td>-0.398 (0.235)*</td>
<td>0.013 (0.042)</td>
<td>0.122 (0.137)</td>
<td>0.014 (0.041)</td>
<td>0.097 (0.112)</td>
</tr>
<tr>
<td><strong>Congressional Neglect</strong></td>
<td>-0.061 (0.043)</td>
<td>0.202 (0.161)</td>
<td>0.071 (0.034)**</td>
<td>0.029 (0.114)</td>
<td>0.097 (0.033)**</td>
<td>-0.012 (0.092)</td>
</tr>
<tr>
<td><strong>Knowledge of GPRA</strong></td>
<td>-0.073 (0.042)*</td>
<td>0.347 (0.198)*</td>
<td>0.054 (0.038)</td>
<td>-0.064 (0.125)</td>
<td>0.063 (0.035)*</td>
<td>0.122 (0.104)</td>
</tr>
<tr>
<td><strong>Number of Employees</strong></td>
<td>-0.146 (0.045)**</td>
<td>-0.705 (0.172)**</td>
<td>0.122 (0.045)**</td>
<td>0.457 (0.119)**</td>
<td>0.116 (0.044)**</td>
<td>0.246 (0.102)**</td>
</tr>
<tr>
<td><strong>Years as Supervisor</strong></td>
<td>—</td>
<td>—</td>
<td>-0.008 (0.041)</td>
<td>0.047 (0.135)</td>
<td>-0.009 (0.041)</td>
<td>0.066 (0.117)</td>
</tr>
<tr>
<td><strong>Grade Level</strong></td>
<td>0.006 (0.057)</td>
<td>-0.142 (0.227)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Involved in GPRA—Developing Measures</strong></td>
<td>-0.017 (0.155)</td>
<td>—</td>
<td>0.474 (0.135)</td>
<td>—</td>
<td>0.569 (0.126)**</td>
<td>—</td>
</tr>
<tr>
<td><strong>Involved in GPRA—Using Measures</strong></td>
<td>0.011 (0.078)</td>
<td>—</td>
<td>0.182 (0.127)</td>
<td>—</td>
<td>0.156 (0.134)</td>
<td>—</td>
</tr>
<tr>
<td><strong>Involved in GPRA—Gathering/Analyzing Data</strong></td>
<td>0.100 (0.093)</td>
<td>—</td>
<td>0.304 (0.127)**</td>
<td>—</td>
<td>0.329 (0.133)**</td>
<td>—</td>
</tr>
<tr>
<td><strong>Headquarters</strong></td>
<td>—</td>
<td>—</td>
<td>-424 (0.099)**</td>
<td>—</td>
<td>-0.306 (0.098)**</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Government Support</strong></td>
<td>0.014 (0.092)</td>
<td>-0.189 (0.836)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Agency Support</strong></td>
<td>0.035 (0.124)</td>
<td>-0.452 (0.385)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Research and Development</strong></td>
<td>0.121 (0.170)</td>
<td>-0.977 (0.705)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Credit</strong></td>
<td>-0.014 (0.012)</td>
<td>2.886 (1.111)**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Grants</strong></td>
<td>0.353 (0.377)</td>
<td>-0.209 (0.587)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Regulation</strong></td>
<td>-4.890 (1.559)**</td>
<td>-1.209 (0.874)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Program—Other</strong></td>
<td>0.038 (0.142)</td>
<td>0.253 (1.227)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Lambda</strong></td>
<td>—</td>
<td>-4.69 (0.138)</td>
<td>—</td>
<td>-3.707 (0.788)</td>
<td>—</td>
<td>-3.329 (0.429)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-3.229 (0.499)**</td>
<td>17.456 (1.81)**</td>
<td>-0.0188 (0.321)</td>
<td>19.342 (1.24)**</td>
<td>-0.0291 (0.310)**</td>
<td>15.010 (0.947)**</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>ρ</strong></td>
<td>-1.00 (0.002)</td>
<td>—</td>
<td>0.676 (0.126)</td>
<td>—</td>
<td>-0.704 (0.078)</td>
<td>—</td>
</tr>
<tr>
<td><strong>Chi-square</strong></td>
<td>131.46**</td>
<td>—</td>
<td>203.57**</td>
<td>—</td>
<td>286.71**</td>
<td>—</td>
</tr>
<tr>
<td><strong>Log-likelihood</strong></td>
<td>-1688.815</td>
<td>—</td>
<td>-5632.771</td>
<td>—</td>
<td>-5365.059</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: Dependent variables: Use and Useful. Estimates from maximum-likelihood Heckman regression model with selection. Robust standard errors reported in parentheses. Base category for Program variables is service delivery.

*p < .10, **p < .05.
a predicted value in the 13th percentile of the Use (2000) distribution. A similar pattern holds across all three dependent variables; Leadership Commitment is an influential factor driving the perceived efficacy of results-model reform and use of performance measures.  

Measurement Challenges, gauging perceptions that technical obstacles to performance measurement would hinder results-model reform, is also a significant factor in the models. With the exception of the Credible (1997) model where the coefficient falls short of significance, Measurement Challenges is strong and negative across the Credible,

---

As noted, an unusually large proportion of responses to the Use (1997), in particular, score a maximum value of 30, leading to reasonable suspicion about the strategic nature of agency manager reporting. However, separate models omitting maximum values in the dependent variables yield similar findings. The Use (1997) model omitting respondents reporting a maximum value in the dependent variable finds Leadership Commitment and Number of Employees significant and positive coefficients and Measurement Challenges and Political Conflict are significant and negative ($N = 498$). The only notable departure from the model reported is the Congressional Neglect variable, which remains positive, becoming marginally significant.
Use, and Useful models. Figure 3b illustrates predicted values for Use (1997, 2000) varying Measurement Challenges. The roughly one-quarter of respondents most concerned about the technical obstacles to performance measurement, again all other factors held constant at their mean values, are associated with levels of Use (2000) in the 22nd percentile among respondents, while for the quarter reporting only moderate concern predict rates of Use (2000) rising to the 35th percentile. By contrast, the two variables gauging obstacles in the political environment—Political Conflict and Congressional Neglect—offer a mixed picture. In 1997, concerns about Political Conflict have a significant negative impact both on the perception that reform is Credible and on reported Use of performance measures. Only 1 year prior, budgetary brinksmanship between congressional Republicans and the Clinton administration led to the December 1995 “government shutdown.” In a political atmosphere dominated by partisan conflict, it is little surprise that managers would believe politics might derail GPRA. In the 2000 models, by contrast, this effect dissipates, generating marginal positive coefficients. In all the models, Congressional Neglect is small and falls well short of statistical significance in the outcome equations.

Measures of existing organizational practices and capacity—Management and Information—yield mixed results. In the Credible (1997) model reported in table 1, Management is positive as expected though falls short of statistical significance. Information is significant and positive as expected in the Credible (1997) model; however, the coefficient reverses signs, taking a significant and negative value in the Credible (2000) model. This pattern also holds in the Use and Useful models reported in table 2. The Information coefficients switch from positive as expected in 1997 to negative in 2000. Though unexpected, this result is plausible if the relationship between Information and the dependent variables operates in both directions. By 2000, when implementation of GPRA had been under way for several years, those managers more committed and confident in GPRA may have developed a new concern about building and maintaining the internal capacity and expertise to support performance measurement. This is an important limitation of the subjective assessments from which these data, and indeed much of the available data in this field, are drawn. The role of shifting expectations in the process of administrative reform

---

23 Again, this variable is omitted from the Use and Useful models due to similarities in the underlying survey items.
is analytically daunting, though it is an essential element in the analysis developed here and is worthy of further development and empirical research.

Measures of respondent characteristics generate a handful of noteworthy findings. Self-reported Knowledge of GPRA switches across the two surveys from positive in Credible (1997) and Use (1997) to negative in Credible (2000) and Use (2000). Of these coefficients, only one is statistically significant, though the pattern is consistent with broad sentiment among observers of agency management, many of whom grew less confident in GPRA’s usefulness during this period in the law’s implementation (GAO 2001). Number of Employees generates positive coefficients across the outcome equations, rising to statistical significance in the Use and Useful models. No firm hypothesis regarding this variable was developed in the construction of the models; however, other factors constant, a positive correlation between number of employees supervised and the reported utility of performance measures is plausible. Grade Level and Years as Supervisor are both positive and significant in the Credible models, suggesting that more senior and more experienced managers report lower levels of confidence in GPRA. In models predicting the Use and Useful dependent variables, however, these variables are mixed and statistically marginal. Finally, the series of dichotomous controls for Program Type included in the Credible models, using service delivery as the base category, yield a handful of significant coefficients though no notable patterns.

Turning to the selection equations reported in the left-hand column under each of the models presented in tables 1 and 2, again two sets of expectations draw on and extend the commitment logic. In the first, NB responses represent uncertainty, which may be reduced with a clear signal—positive or negative—regarding the credibility of leadership commitments. NB responses are therefore minimized at both high and low values of Leadership Commitment, forming a curvilinear pattern; thus, a directional hypothesis was not specified. The second expectation follows from the proposition that NB responses are a function in part of survey satisficing. From this perspective, credible Leadership Commitment should increase manager attentiveness, and thus, the likelihood managers will offer substantive, as opposed to NB, responses. In tables 1 and 2, however, Leadership Commitment is negative and significant in the Credible (1997) and Use (1997) selection equations. Managers reporting committed agency leadership were significantly less likely to offer a response when asked whether they use performance measures or believe GPRA will help improve agency management. In the 1997 models, highly committed agency leaders are apparently associated with more rather than less uncertainty when it comes to reform’s consequences. In the 2000 Use and Useful selection models, by contrast, Leadership Commitment is significant and positive. Figures 4a and b illustrate these divergent results, graphing the predicted likelihood of a response in two models—Credible (1997) and Use (2000)—based on different values of Leadership Commitment, holding other variables at their means. At the minimum value of Leadership Commitment, both models generate a predicted value of about \( p = .86 \). As Leadership Commitment rises from minimum to maximum values, the predicted likelihood of a response to the Credible (1997) variable drops from \( p = .86 \) to \( p = .69 \), whereas the predicted likelihood of a response to the Use (2000) dependent variable increases from \( p = .86 \) to \( p = .95 \).

It is instructive to compare these divergent results with the other coefficients in the selection equations as other negative perceptions also appear to factor prominently in respondents’ decisiveness. Political Conflict is significant and positive in the Credible (1997) selection equation, remaining positive though short of significance in all but one
of the models. In other words, managers who fear political conflict will hinder the implementation of GPRA are more likely to offer an opinion regarding its prospects. Likewise, Measurement Challenges is positive and significant in Use (1997) and Credible (2000) selection models and Congressional Neglect is positive and significant in the Use (2000) and Useful (2000) selection models. Though not uniformly robust, the significant factors in the selection equations reinforce this pattern. The findings tend to support an interpretation of NB responses rooted in skepticism, where factors hindering the prospects of reform weigh heavily in the formation of respondents’ expectations. The finding that low levels of Leadership Commitment are associated with an increasing likelihood of a response in the Credible (1997) and Use (1997) models conforms to this pattern. Managers tend to weigh problems more heavily than potential when it comes to reform and weak Leadership Commitment signals reform’s prospects are weak. In the Use (2000) and Useful (2000) models, however, the coefficient reverses. Increasing Leadership Commitment is positively associated with likelihood of a response.\textsuperscript{24} Here, a plausible case can be made that Leadership Commitment reduces uncertainty, signaling reform may indeed take hold. At a minimum, the relationship between Leadership Commitment and the dependent variables tests the limitations of the relatively simple model developed here. The dual interpretation above suggests a curvilinear relationship. For current purposes, it is sufficient to observe that across models Leadership Commitment is a vital factor framing the prospects of reform.\textsuperscript{25}

\textsuperscript{24} Other variables in the selection equations are mixed but conform generally to expectations. Information, measuring perceived organizational capacity and expertise, positively predicts the likelihood of responses in the 2000 Use and Useful models. Self-reported Knowledge of GPRA positively predicts the likelihood of a response to the Credible questions in 1997 and 2000 and the Useful questions in 2000, though this variable falls short of significance for Use (2000) and is marginally significant and negative in the Use (1997) model. In the Credible models, both Grade Level and Years as Supervisor are significant and positive and Number of Employees is not significant. In the Use and Useful models, on the other hand, neither Grade Level nor Years as Supervisor are significant; but Number of Employees is significant, switching from negative in Use (1997) to positive in Use (2000) and Useful (2000).

\textsuperscript{25} A recent study of perceptions and use of performance measures in local government agencies yields a similar result using a “lasting effect” dependent variable (Melkers and Willoughby 2005).
Evidence presented here and elsewhere suggests leadership commitment is a critical factor shaping the perceived impact of management reforms, but the concept remains frustratingly difficult to observe and analyze in a systematic way. This article contributes to a growing literature examining the implementation of contemporary results-model reforms by clarifying the concept of leadership commitment in two ways. The first section borrows the logic of credible commitment from institutional economics, reasoning that reform often fails under the burden of the vertical dilemma characterizing relations between principals and agents (Miller 1992). Would be reformers face a credibility problem: Are reform commitments simply cheap talk? If managers suspect agency leaders are prone or simply will not be around to make good on reform commitments, their attention and effort will gravitate to other problems and priorities. The nature of the public sector—and of many types of organization—is such that the fabric of leadership commitments necessarily relies on reputational credibility. Reputation-based credibility is more time consuming to cultivate and difficult to maintain than are formal, regulative commitments. However, normative professional commitments like the politics/administration dichotomy—which despite a tumultuous intellectual lineage remains a widely shared “useful myth” in many public sector organizations—lend a vital resource from which to construct a reputation for good behavior (Miller 2000).

The article’s second section presents a quantitative analysis drawing data from two GAO surveys of federal agency managers. Administered in 1996–97 and 2000, during the implementation of the GPRA, these data highlight the central importance of leadership commitment for both the credibility of results-model reforms and the reported use of performance measures across government agencies. As expected, the empirical analysis shows Leadership Commitment is a robust predictor of the three dependent variables Credible, Use, and Useful. Managers who do not believe an agency’s leadership is committed to performance measurement are less likely to view GPRA’s reform mandates as credible, less likely to report using performance measures, and less likely to report they find performance measures useful in accomplishing managerial tasks. Among other expected effects, Measurement Challenges is a prominent, negative factor across models and Political Conflict is significant in the 1997 models.

The Heckman model applied here to control for the large proportion of managers responding “no basis” or NB to the questions underlying the Credible and to a lesser extent the Use and Useful dependent variables suggests that leadership commitment influences not only substantive evaluations of reform but also the likelihood managers even offered an opinion. Across the 1997 models’ most prominent indicators—Leadership Commitment, Political Conflict, Measurement Challenges—pessimistic attitudes make managers more rather less likely to provide a substantive response. Likewise in the 2000 models, concerns about Congressional Neglect increase the likelihood of a response; however, with the exception of one nonsignificant variable, the Leadership Commitment coefficient reverses, instead becoming significant and positive. In the 2000 models, positive assessments about Leadership Commitment increase the tendency of managers to offer an evaluation. Though the analysis offers only limited basis for inference, one speculation is that the relationship between Leadership Commitment and uncertainty is curvilinear, with uncertainty decreasing at both the high and low extremes of the Leadership Commitment spectrum.
At low levels of leadership commitment, managers may understandably conclude the demands of maintaining the status quo outweighs the less than tangible promise of reform. When presented with uncertain trade-offs, borrowing an image from decision theory, losses loom larger than gains (Kahneman and Tversky 1979; March 1994). If managers exhibit skepticism in their judgments of contemporary results-model reform, the social and political context only reinforce this disposition. In a contemporary study of executive-level agency politics, Aberbach and Rockman observe, “GPRA leaves it to the agencies to propose goals. That can put them in jeopardy with congressional committees and constituencies.” They conclude, “It is hardly surprising that agencies have often shied away from stepping on this potential landmine” (Aberbach and Rockman 2000, 150–1). Behn evaluates the performance bargain—promising results-based accountability for enhanced managerial discretion—and concludes, “To the distrustful (or acute, or pragmatic, or cynical) public manager, the big bargain of performance management is no bargain at all” (Behn 2002, 9, 15). Performance measurement takes time and poses risks for managers that, absent evidence the other half of the performance bargain will materialize, may look like a losing proposition. At the very least, these initiatives cannot have been expected to inspire vigilance.26

Consider the comments of two field office managers from the Social Security Administration (SSA) speaking during one of a series of small “peer group” interviews conducted with agency managers in conjunction with this research project.27

But, despite what I’ve said here, I tend to actually be pretty positive about change... My usual reaction is that if that’s what they want us to do, I’m going to try to do it the right way. I’m going to try not to fake it. But, as each change occurs I do sort of appreciate the fact that I’m past 55 and I’ve got more than 30 years of service. (SSA 1.1)

My first reaction is always fear. Fear that it’s going to be punitive or counterproductive to how I’m used to doing things. And, fear borne out of the fact that oftentimes these initiatives are formulated at such a distant level that it really doesn’t take into account local needs, whether that’s a field office or a larger operation. It seems like it’s a disconnected process from what we do at a local level serving the public. (SSA 1.2)

SSA is by most accounts one of government’s performance management success stories, though these two SSA managers evince no uniform commitment to reform. The first (SSA 1.1), a district manager who oversees a field office in a growing Southeastern community, notes that he has been an SSA employee for more than 30 years. Throughout the discussion, he spoke jokingly about the various management initiatives he has seen come and go. His comments above, which drew chuckles from other participants, reassured the group that in spite of his lighthearted cynicism he remained committed to improving management. The comment elicited a like-minded response by a district manager from the

26 Boyne’s (2004) study of the implementation of “best value” model planning by local authorities in the United Kingdom offers a contrasting finding.
27 Six peer group discussions were conducted with managers in two government agencies, the SSA and the Internal Revenue Service. Participants in the peer groups were drawn from the same population of mid- and upper-level career managers the GAO surveys drew from. These small three- to five-person group discussions were conducted via conference call and were between 90 and 120 min in length. A detailed methodology and review of peer group findings are included in Dull (2006a).
Midwest, “I’d like to take a swing at that too...I guess I try not to pre-judge these things. And, I would start by saying that some of the change I see in the agency now I regard as being among the most positive I’ve ever seen. For my money the commissioner we have now is probably the best we’ve had since I’ve worked here” (SSA 1.3). This positive assessment of Commissioner Jo Anne Barnhart’s leadership brought assent from other participants in the conversation, including the participant responsible for the second, more pessimistic comment above. This district manager from a rural area in another midwestern state responded, “One thing that makes the commissioner and deputy commissioner different from past administrators is their willingness to listen and actively seek feedback. We haven’t always had that. That’s greatly appreciated by me and I think it impacts their decisions as well” (SSA 1.2). His apparent respect for the Deputy Commissioner for Operations Linda McMahon, a point he returned to, stands in contrast to his general suspicion of senior level SSA administrators and their reform commitments. Whereas the first manager slyly suggests he can afford not to worry about the consequences because, “I’ve got more than 30 years of service,” in other words, he can retire; the latter expresses fear that initiatives will be “punitive,” and the “distant” and “disconnected process” leading to their implementation will disrupt his working “at a local level serving the public.”

Much has been said and written about the reform idea as it relates to public sector governance, much of it embracing a well-earned skepticism about the realism of reform promises. Yet, reform—represented here by the results-model, emphasizing the use of strategic planning and performance measures—remains a vital feature of public sector leadership (Dull 2006b; Frederickson and Frederickson 2006). The analysis presented here examines reform from a distinct perspective, that of the middle- and upper-level agency managers who form a critical link between reform ideas and organizational performance. As the SSA managers quoted above illustrate, responses to questions of leadership reform commitments are diverse, but they echo a common wariness. Whatever the particulars of reform leadership, credibility remains an active ingredient.

**APPENDIX: SURVEY DATA AND VARIABLES**

**GAO Survey Methodology**

Data are drawn from two surveys administered by the GAO to samples of mid- and upper-level managers and supervisors in the 24 executive branch agencies covered by the Chief Financial Officers Act, a population that accounts for approximately 97% of the executive branch full-time workforce. Using the Office of Personnel Management’s Civilian Personnel Data File, GAO identified managers (employees accountable for the activities of programs/operations/projects) and supervisors (employees responsible for the work of others) of pay grade level GS-13 to GS-15/SES. The first survey was administered to a stratified random sample of 1,300 managers and supervisors between November 27, 1996, and January 3, 1997, resulting in 905 useable survey responses. The second survey was administered to a stratified random sample of 3,816 managers and supervisors between January and August, 2000, resulting in 2,508 useable survey responses. Though the survey instrument was revised and expanded between the first and second surveys, unless otherwise indicated, all survey items identified below are identical. Unless otherwise indicated, all items employ five-point, closed-ended scales. Additional details regarding survey methodology are provided in GAO reports (GAO 1997, 2001).
Variables

Credible
Additive index of two survey items:

− “To what extent, if at all, do you believe that your agency’s efforts to implement GPRA to date have improved your agency’s programs/operations/projects?”

− “To what extent, if at all, do you believe that implementing GPRA can improve your agency’s programs/operations/projects in the future?”

Use
Additive index combining six items under a single question: “For those programs/operations/projects that you are involved with, to what extent, if at all, do you use the information obtained from performance measurement when participating in the following activities?”

− “Setting program priorities.”

− “Allocating resources.”

− “Adopting new program approaches or changing work processes.”

− “Refining program performance measures.”

− “Setting individual job expectations for the government employees I manage or supervise.”

− “Rewarding government employees I manage or supervise.”

Useful (2000 only)
Additive index combining five items under a single question: “For the performance measures you have, to what extent, if at all, do you find them useful for managing the programs/operations/projects that you are involved with?”

− “Performance measures that tell us how many things we produce or services we provide (output measures).”

− “Performance measures that tell us if we are operating efficiently (efficiency measures).”

− “Performance measures that tell us about the quality of the products or services we provide (quality measures).”

− “Performance measures that tell us about the quality of the products or services we provide (quality measures).”

− “Performance measures that would demonstrate to someone outside of our agency whether or not we are achieving our intended results (outcome measures).”

Leadership Commitment
Single survey item:

− “Extent to which I agree with the following statement: Currently, my agency’s top leadership demonstrates a strong commitment to achieving results.”
Management Accountability and Recognition
Average of two survey items: “Extent to which I agree with the following statements. . .”

  – “Currently, agency managers/supervisor at my level are held accountable for the results of the programs/operations/projects they are responsible for.”
  – “Currently, employees in my agency receive positive recognition for helping the agency accomplish its strategic goals.”

Information Capacity and Expertise
Higher value of two items: “Based on your experience with the programs/operations/projects that you are involved with, to what extent, if at all, have the following factors hindered measuring performance or using performance information.”

  – Existing information technology and/or systems not capable of providing needed data.
  – Lack of staff who are knowledgeable about gathering and/or analyzing performance information. Scale has been reversed so that higher values are associated with higher perceived levels of information capacity and staff expertise.

Political Conflict (1997)
Highest value out of four items: “Based on your experience with the programs/operations/projects that you are involved with, to what extent, if at all, have the following factors hindered measuring performance or using performance information. . .”

  – “Difficulty resolving conflicting interests of stakeholders, either internal or external.”
  – “Concern that performance information could be used against my program or agency.”
  – “In your opinion, how easy or difficult will it be (or has it been) for your agency to reconcile differing congressional views on the agency’s mission and strategic goals?”
  – “In your opinion, how easy or difficult will it be (or has it been) for your agency to reconcile the views of other interested parties on the agency’s mission and strategic goals?”

Political Conflict (2000)
Two items included in the 1997 survey are not available in the 2000 survey. Takes the higher of two items: “Based on your experience with the programs/operations/projects that you are involved with, to what extent, if at all, have the following factors hindered measuring performance or using performance information. . .”

  – “Difficulty resolving conflicting interests of stakeholders, either internal or external.”
  – “In your opinion, how easy or difficult will it be (or has it been) for your agency to reconcile the views of other interested parties on the agency’s mission and strategic goals?”

Congressional Neglect
A single survey item: “Based on your experience with the programs/operations/projects that you are involved with, to what extent, if at all, have the following factors hindered measuring performance or using performance information.”
–“Lack of ongoing congressional commitment or support for using performance information to make program/funding decisions.”

**Measurement Challenges (Linking Inputs to Outcomes)**

Highest value of four survey items: “Based on your experience with the programs/operations/projects that you are involved with, to what extent, if at all, have the following factors hindered measuring performance or using performance information.”

–“Difficulty obtaining valid or reliable data.”

–“Difficulty obtaining data in time to be useful.”

–“Results of our program/operation/project occurring too far in the future to be measured.”

–“Difficulty distinguishing between the results produced by the program and results caused by other factors.”

**Knowledge of GPRA**

A single item: “Prior to receiving this questionnaire, which of the following statements best describes your awareness of GPRA?”

1. I had never heard of GPRA.
2. I had heard of GPRA but had no knowledge of its requirements.
3. I had heard of GPRA and had a low level of knowledge of its requirements.
4. I had heard of GPRA and had moderate knowledge of its requirements.
5. I had heard of GPRA and had extensive knowledge of its requirements.

**Number of Employees**

A single survey item: “In your current role, approximately how many government employees are you responsible for?”

1. 0–10
2. 11–50
3. 51–100
4. 101–300
5. >300

**Years as Supervisor (2000)**

This variable is based on an open-ended question, scripted, “In total, for how many years have you been a supervisor and/or a manager in the federal government.” GAO summarized responses in the following ordinal scale:

1. <5 years
2. 6–10 years
3. 11–19 years
4. >20 years
**Grade Level (1997)**
A single item: “What is your current grade level?”

1. GS/GM-13 or equivalent
2. GS/GM-14 or equivalent
3. GS/GM-15 or equivalent
4. SES

**Involvement in GPRA (Developing Measures)**
A single “Yes/No” item: “During the past 3 years, have you been directly involved in any of the following GPRA-related activities?”

—“Developing ways to measure whether program performance goals are being achieved.”

**Involvement in GPRA (Gathering/Analyzing Data)**
A single Yes/No item: “During the past 3 years, have you been directly involved in any of the following GPRA-related activities?”

—“Gathering and analyzing data to measure whether programs are meeting their specific performance goals.”

**Involvement in GPRA (Using Measures to Track Goals)**
A single Yes/No item: “During the past 3 years, have you been directly involved in any of the following GPRA-related activities?”

—“Using measures for program goals to determine if the agency’s strategic goals are being achieved.”

Dichotomous variable based a single item: “Please indicate where you currently work. (If you are currently on temporary assignment or on detail please answer for your permanent work location. Check one.)”

—“Headquarters of my department or agency.”

—“A field office of my department or agency.”

—“Other—please specify.”

The latter two response categories are collapsed. Headquarters = 1, all others = 0.

**Defense (1997)**
A single Yes/No item: “Are you employed by a defense or a nondefense agency?”

**Program Type (1997)**
Series of dichotomous variables from a single item with the following descriptions of each function type:

—Federal government–wide support—“Direct services to federal employees or agencies, such as human resources management by the Office of Personnel Management or procurement by the General Services Administration.”
–Internal agency support—“Staff offices such as planning, policy, human resources management, procurement, fiscal management, or Office of the Secretary.”

–Research and development—“Research to increase technical knowledge or develop new or improved products.”

–Credit—“Monetary credit provided to citizens or nonfederal entities through direct loan or loan guarantee.”

–Grants—“Formula-based grants or those that provide funding for specific activities or projects in response to applications or proposals by individuals or institutions.”

–Regulatory/enforcement—“Develop, promulgate, and enforce rules and procedures governing the behavior and operations of individuals, firms, organizations, agencies or communities.”

–Service delivery—“Provide direct services to citizens or nonfederal entities.”

–Other.

In the accompanying quantitative models, the omitted variable is service delivery, a program-type often considered amenable to performance measurement.

**FUNDING**

The IBM Center for the Business of Government, Washington, DC.

**REFERENCES**


