From Measurement to Management: Breaking through the Barriers to State and Local Performance

Mary Bryna Sanger
Milano: The New School for Management and Urban Policy

The Winter Commission’s aspirations for improvement of governance at the state and local level resulted in proposals to modernize public institutions and systems to remove the barriers on executive and managerial authority to act. Improving government performance and accountability to citizens requires leadership to empower employees by reducing rules, increasing discretion, and rewarding innovation. The performance measurement movement and its related performance management movement are public management trends of wide influence in state and local government that are both an adjunct to, and a reflection of those aspirations. Case studies of cities that have sought to develop and use performance measurement and engage citizens demonstrate great potential to improve governance and manage for results that citizens want. But even exemplary jurisdictions face challenges using evidence-based, data-driven performance management approaches. Promoting best practice requires better scholarship to understand the determinants of successful implementation, adoption, and use of performance measurement. The outcomes of performance management systems are generally unmeasured and little is known about their cost effectiveness or endurance over time. The promise is there, but our expectations should be tempered.

The Winter Commission’s proposals 15 years ago sought to create more efficient, accountable government that would produce results citizens want. The aspirations at the heart of many of the commission’s proposals were to remove barriers to executive and managerial authority to act—what was dubbed the “trust and lead” approach. The various proposals sought to provide executives and their managers the flexibility and freedom to change and innovate and to recruit, hire, and develop knowledgeable and motivated employees who could be held accountable. Further, the proposals recognized the need to engage citizens and enlist them (Thompson 1993).

Freeing executives and managers to act and arming them with evidence-based, data-driven decision tools promotes performance-managed government. The performance measurement movement is an adjunct and a result of that effort. A recent review of public management efforts around the country identified performance measurement as a trend transforming government because of its potential to improve government performance and accountability (Abramson, Breul, and Kamensky 2006). How much the movement has changed the behavior of state and local governments and how well citizens are engaged in either its development or its consumption are important questions about the role performance measurement has played in improving governance. Further, how much these developments have leveraged significant improvements in public management that are at the heart of the commission’s aspirations are the key questions addressed in this article.

How much has actually been accomplished is a key research question. To what degree are governments measuring and sharing the results citizens want, and how are they determining what those are? Further, to what degree do governments and agencies actually realize the potential of measurement for managing their operations, motivating and engaging their workforces, and allocating their resources? These questions are central to the commission’s aspirations, and their answers provide an agenda for continued research and reform.

Progress is evident in efforts around the country, but it is uneven. More meaningful efforts are under way in cities than in states, and agencies use performance management more often and effectively than jurisdictions (Moynihan 2008). This article reviews some recent efforts at performance measurement in state and local governments around the country (especially those that engage citizens) and considers their role in achieving greater accountability, responsiveness, and operational performance of public agencies. Through several additional case studies examining cities, states, and agencies that have attempted to use performance measures to
drive management improvement, the article will further examine the leading obstacles to performance improvement through measurement and management efforts. Finally, knowledge deficits impede our ability to conclude with confidence what explains much of the variation we see. Throughout, we suggest the shortcomings of existing research and where future study might help to fulfill the promise of improved performance that accompanied the commission’s original aspirations.

**Why Measure Performance?**
Achieving the commission’s goals depends on executives and managers having authority, but it also relies on the translation of their goals and objectives into concrete and measurable results. There is broad agreement in theory on the numerous ways performance measurement can help, even while there is significant debate about its failings in practice (Moynihan 2008; Radin 2006; Schick 2001). Performance measurement has been proposed in different guises since the Progressive Era (Ridley and Simon 1938), and waves of reform over the last century have alternatively emphasized economy, efficiency, waste and fraud, and transparency (Light 1997). The value of performance measurement is recognized for helping executives hold their managers accountable for meeting their mandates and for promoting useful reward, sanction, and motivational efforts to improve the performance of management down the line. The information it produces permits executives to judge the success of their operations and their improvement over time, allowing them to tailor new strategies to changing needs or priorities (Behn 2003). Performance measurement can also help executives and managers save resources, allocate them more effectively, and plan future investments in changing budgetary environments (Ammons 1995; Behn 2003). Not everyone agrees on the best measures to use or how heavily to weigh them in light of competing goals, but performance measurement is a train that left the station a long time ago.¹

**Manager and Stakeholder Value**
While evidence indicates that performance measurement is most often valued for program management (Ho 2003; Lynch and Day 1996), there is also a growing appreciation that performance measurement has significant value to all stakeholders, including citizens and legislators. Politicians and elected officials can learn how well services are delivered, how well they meet the needs of their constituents, and whether they reflect the political demands they are elected to fulfill. This knowledge is critical for governing, as it helps to maximize politically valuable outcomes in contested environments.

Performance measurement and its public reporting . . . have the capability of restoring citizen trust in government by making its activities—service efforts and accomplishments—more transparent, open to public scrutiny, and demonstrative of real value for taxpayers. Performance measurement, therefore, can build political and civic support for public efforts and increase the legitimacy of public authority (Ho 2003). Thus, performance measurement that is widely available, reflective of citizens’ interests, and subject to feedback through available institutional structures can strengthen local democratic institutions and make local government more accountable to citizen demands. In this way, citizens are seen not as customers alone but, as Ho and Coates point out, as “owners, issue framers, coproducers, and evaluators of government” (2004, 29).

**Monitoring Recent Performance Measurement and Its Use in Cities and States**
Most of our knowledge about what states and cities have been doing since the Winter Commission comes from three sources: selected case studies by organizations that sponsor and promote the use of performance measurement or service efforts and accomplishments (SEA) measurement in state and local government;² surveys by scholars or sponsoring organizations of state and local government officials about the use and value of metrics (Berman and Wang 2000; GASB 2002a; Holzer and Kloby 2005; Melkers and Willoughby 2000, 2004, 2005; Poister and Streib 1999); and document reviews and interviews with officials (Moynihan 2001; Smith et al. 2007). As a result, the bulk of the research findings on the nature and use of performance measures rely most often on self-reported information by state and local officials.
who are under pressure from multiple constituents to demonstrate their support for and compliance with behaviors that increasingly are becoming professional norms. Case studies provide more depth and reliability but, of course, do not reflect the breadth of efforts around the country.

Our knowledge about performance measurement efforts at the local level have similar sources: surveys of government officials, document reviews, and case studies. But because the cities and counties employing these tools are so numerous (about 87,000 at last count), researchers have typically limited themselves to local jurisdictions of certain sizes or characteristics, and response rates have been far lower than those for states (Ingraham, Joyce, and Donahue 2003; Ho and Ya Ni 2005). Survey results can be difficult to interpret because local government services exhibit wider variation in the services provided and the sources of provision (city, county or special district) than do states. Thus, they are hard to compare. Like states, research relies on self-reports and surveys that are not generally comparable from researcher to researcher (Berman and Wang 2000; De Lancer Julnes and Holzer 2001; Garsombke and Schrad 1999; GASB 2002a; Ho and Ya Ni 2005; Ingraham, Joyce, and Donahue 2003; Poister and Streib 1999). Caution is required when accepting these reports of performance measurement on their face, as examinations of actual documents reveal exaggerated claims of actual use (Ammons 1995).

**States**

Performance measurement in the states has been highly touted, enabled by improved technology, and increasingly required by law or administrative mandates. When viewed over time since the commission made its proposals, most states report that they measure performance and manage for results; follow-up research shows that measurement is evolving and enduring (Melkers and Willoughby 2000, 2004).

Neither legislative nor administrative requirements in the states, however, necessarily call for the application of performance measures, nor do most state agency goals, however measured, to the budget process in all agencies (Melkers and Willoughby 2004). Further, comprehensive performance or SEA data in the states are not often easily available to the public or posted on Web sites. When they are, the data are frequently out of date (Smith et al. 2007). Further, citizens are rarely consulted or involved in their development.

The quality, character, breadth, and value of performance measurement in the states and its relationship to the outcomes citizens want and managers need for organizational decision making are not sufficiently documented. Whatever the promise of the performance movement, the quality and value of states’ adoption is highly variable, and the impact of the measures on management outcomes is relatively unmeasured (Ingraham, Joyce, and Donahue 2003). Further, there is often controversy about the appropriateness and impact of the measures selected, leading critics to point out the significant opportunities for perversion, distortion, and unintended consequences (Radin 2006; Smith 1995). Even so, as we will discuss in a following section, good performance measurement is essential, if not sufficient, for performance management. And while state capacity to manage for results has been improving, it remains inconsistent and, in many cases, unreliable. Few states get good grades overall for the quality of the information they collect, share, and use for accountability and decision making even in the most recent comprehensive evaluations (Barrett and Greene 2008; Moynihan 2001). Only political commitment and sustained leadership has been consistently found to be related to achieving high levels of performance management capacity, if not results (Behn 2004; Moynihan and Ingraham 2003; Sanger 2008).

**Cities and Counties**

We would expect cities and counties to do better in measuring performance, consulting with and reporting to citizens about it, and using it to manage. Cities and counties provide citizens more visible services that they understand and that impact their daily lives. Their measurement is technically easier than it is for many state functions, and citizens are more likely to hold local officials accountable for direct service delivery. Not surprisingly, therefore, in the last 15 years, cities and counties around the country have embraced performance measurement for a greater range of services and with more application to performance management. They represent diverse jurisdictions with varying economic capacity, size, institutional structures, and political cultures.

With all the caveats associated with the data, we know that budget officials, auditors and agency staff in more and more local jurisdictions are reporting that they collect performance data, with increasing numbers reported by various surveys over the decade. While input and workload data are those most consistently reported, increasing numbers of jurisdictions are using output measures—those most important for managing for performance (GASB 2002a; Ho and Ya Ni 2005).

In contrast to states, according to a 2002 Governmental Accounting Standards Board (GASB) survey, 43 percent of city and county respondents reported that their jurisdiction used quality or customer satisfaction data. Further, the GASB’s survey found that more than 64 percent of city and county respondents reported that at least 50 percent of agencies had output or outcome data appearing in annual operating budgets, and 58 percent had them in budget requests. Despite the increasing numbers of jurisdictions that claim to be...
collecting and using performance measures, there is still some question about how much local agency operations are actually driven by them and how deeply managing for results has actually taken hold (GASB 2002a).

Some indications from the Government Performance Project 2000 grades of the 35 largest cities suggest that when comparing cities to states in managing for results, cities do better (Moynihan and Ingraham 2003). This is not surprising. Cities have historically been leaders in the performance movement. According to the study, cities not only attained better grades than states but also displayed a wider range of strategies. While the challenge for defining their goals and service outcomes is simpler, they have been shown to be making good progress in understanding the utility of various means and levels of measurement and in crafting appropriate measures for different audiences and targets (Ingraham, Joyce, and Donahue 2003).

Citizens, however, are consistently the least mentioned impetus for the development of performance measurement and management systems (Garsombke and Schrad 1999; GASB 2002a; Melkers and Willoughby 2005). Only 13 percent of city and county respondents in a recent GASB survey involved citizens in the selection of performance measures (GASB 2002a), and only 36 percent believed that citizens, clients, or stakeholders were very important in performance measurement (GASB 2002a). Fewer than a third of cities and counties reported that they distribute and prepare performance measures for public reporting to citizens.

Performance measures for both state and local governments are primarily performed for the mayor, the governor’s office, or the city manager, according to multiple surveys of governmental officials at both the state and local level. Reporting outcomes has more meaning and value to citizens, but it poses more political risks to agency directors and managers (Ho and Ya Ni 2005; Ingraham, Joyce, and Donahue 2003). And while executives and managers need performance information to manage, they often fear being held accountable for results they believe they cannot entirely control. Thus, while much has been demonstrated about the importance of having the technical capacity and professional expertise for measuring and using measures for performance management, political support and leadership at the executive and managerial level matter as much for cities and counties as for states (Melkers and Willoughby 2005).

How performance measurement efforts and reporting to citizens develop and survive at the local level has been a neglected area of research given the importance it has for improving governance. In response, we did six case studies of cities’ efforts to measure performance, seeking both to evaluate their engagement with citizens and their use for management. In a second set of cases, we sought to capture the range of performance management approaches that rely on measurement and how they are applied in varying institutional contexts—jurisdiction-wide and at the agency level. What makes this work different is its ability to test many of the prevailing hypotheses while observing the significance of particular contextual variables not well explored in the literature. The remainder of the article will discuss the findings from studies of both citizen engagement and performance management.

A Closer Look at Six Cities
We undertook a “reconnaissance mission” to a small group of carefully selected jurisdictions to provide some preliminary conclusions about what types of local governments are most likely to undertake these kinds of efforts; the environmental, cultural, and political factors affecting their success; and the factors relating to their use and sustainability. Further, because performance measurement has such potential to improve civic engagement and government accountability, the objective was to identify, where possible, effective means to engage citizens. It is an initial effort to reflect more generally about significant factors that appear to be supporting or impeding successful efforts in a few select cities. The concluding section will look at a few ambitious efforts to use performance measures to manage at the state, local, and agency levels. Our findings demonstrate that even jurisdictions with significant predisposing characteristics for success face very serious challenges. More promise appears to be evident at the agency level.

We began by identifying cities around the country that have undertaken or demonstrated some effort to launch significant performance measurement systems. We were also interested in cities attempting potentially innovative strategies to engage citizens, use performance measurement, and connect the two. These we found through generally published listings of jurisdictions participating in any number of efforts to build and use performance measurement, benchmarking, or community indicators. We included for our initial considerations places where we could collect relevant data on efforts quickly from organizations or sponsors. We identified six jurisdictions in 2004 that were a good mix of the more successful efforts around the country. Each varied with respect to the maturity of their efforts, the sophistication of their measures, the depth and level of citizen engagement, the politics of performance measurement, and the degree to which performance data was used systematically to manage. We selected cities from different regions of the country. The cities ranged in size from 1.3 million (Phoenix) to 172,000 (Worcester). We selected a range of cities a priori: cities with reputations for exemplary...
performance measurement and reporting systems (Portland, Oregon, and Phoenix), innovative means to engage citizens, (Des Moines and Worcester), and comprehensive performance management systems\(^6\) (Baltimore). We also chose San Diego, a hybrid with a somewhat comprehensive performance measurement system of moderate depth (more outputs than outcomes) and a meaningful effort to engage citizens through focus groups in the initial development.

Our data derive from semistructured interviews with actors in each jurisdiction. Over three to four days, we conducted 8–12 interviews with representatives of each city government (elected and management), community and citizen representatives, civic and business leaders, and a few academics and consultants. Supplemented by reports and published documents, we sought to learn the history of the efforts, the role of various players, and the use, impact, and satisfaction with existing performance measurement systems. These interviews were conducted in late 2004. In 2008, we tracked changes and endurance in the efforts over time using telephone interviews, literature, and document reviews.

Our findings paint a picture of the impediments and challenges facing efforts to increase the use of performance measurement, increase its value for management, and increase citizen engagement. While some of these efforts are young and the findings represent the experiences of only six jurisdictions, they are illustrative and may help to illuminate where performance measurement and management reform could strengthen the effectiveness and democratization of public administration consistent with many of the Winter Commission’s prescriptions.

**What Can We Learn from Exemplary Cities?**

We expected better and more comprehensive performance measurement efforts in cities with professional management and high levels of social capital, and in progressive and socially hospitable environments with high levels citizen trust (Sanger 2004). For cities with less than comprehensive systems, we expected measurement efforts to reflect citizen priorities—concrete and observable neighborhood services such as trash collection, graffiti removal, or street conditions that influence quality of life everyday (Berry, Portney, and Thomson 1993; Kelly 2003).

No one characteristic or set of circumstances explains the embrace of performance measurement or the nature of citizen engagement in all our study sites. Each jurisdiction has its own history, politics, and civic infrastructure, and no single element appears to be salient across sites. For example, Phoenix and Portland are cities that are well known for their transparency and civic engagement, and they have been building and refining performance measurement systems for at least a decade (see table 1).
They represented exemplary systems in 2004 in terms of the quality of their reporting and the comprehensiveness of their measures. Both benchmark against comparable cities, have been studied by the GASB, and publish annual performance measurement documents, which they distribute broadly and post on their Web site (GASB 2002b, 2004). But in practice, even exemplary systems are imperfect. Their reliance on performance measurement for management varies across departments; outcomes are not generally reported on a neighborhood basis (a critical value for citizens and community groups); and neither uses performance-based budgeting, even though agency performance measures are often included in the budget reports.

A strong champion within city government appeared to be essential in all our cases, but the champion was equally likely to be a city manager as a mayor—or, in the case of Portland, an independently elected auditor. Not surprisingly, champions with greater power, resource commitments, capacity, and expertise can ensure a higher level and breadth of investment across the board. Phoenix’s city manager, Frank Fairbanks, and Portland’s mayor, Vera Katz (and Mayor Neil Goldschmit before her) were early champions, willing and able to devote considerable political, financial and managerial resources to the task (GASB 2002b, 2004). But success has varied over time, and not all the criteria associated with exemplary systems have been met. Our interviews confirmed that Portland’s citizen engagement is limited to satisfaction surveys, and the SEA reports are not well known by all major constituents. Phoenix does little to systematically use what it collects for budgeting or management and does not engage citizens in ongoing development of the measures.

Des Moines enjoys a culture that values transparency and a city manager who appreciates the returns from a data-driven approach to managing. But he takes a subtle approach to driving performance measurement across service areas, and, like managers in most of our cities, he is cautious about using its results to hold managers accountable. Cost, time, and capacity constraints have also compromised its broad and uniform application. Budget constraints in the early years of development limited the ability to build a comprehensive system across all service areas. But the most recent report demonstrates greater scope, even though few measures reflect on outcomes or results. Instead, they are input or workload measures related to broad strategic goals (such as reporting on the number of fire inspections to reflect on the strategic goal of safety). While capacity and expertise were in short supply during the development stage in 2004, there was also evidence of a philosophical problem in some service divisions. “Some never see the benefit and there has been a failure of communication to demonstrate effectiveness,” reported Mike Matthes, the assistant city manager, whose portfolio includes performance measurement.

Despite such cultural resistance in the government, Des Moines has been ambitious in its outreach to its citizens. The city conducts citizen satisfaction surveys and has worked with communities through a university collaboration to identify measures of interest in local neighborhoods. Citizens were then trained in the use of handheld computers (ComNET) to measure neighborhood conditions annually on valued outcomes. The citizens’ observations and priorities were communicated to city agency computer systems, and the workload data were used to reorder the service priorities of city agencies.

In Baltimore, CitiStat, a comprehensive performance management system implemented across all city agencies, was imposed from the top down by a strong mayor in 2000. Committed to Baltimore neighborhoods from which his political strength derived, Mayor Martin O’Malley (now Governor O’Malley) built a culture and a management apparatus modeled after CompStat, the highly publicized system of performance management used by the New York City Police Department under Mayor Rudolph Giuliani in the 1990s. Dependent on continuous, accurate, and timely information, reported on a geographic basis, CitiStat built a management information system based on continual accountability and reporting all the way down the line. From senior managers to service workers on the ground, CitiStat performance data allow for rapid deployment of resources where they are needed in neighborhoods within the city. Formal biweekly public reviews by the mayor and his senior CitiStat staff of department heads and their staff, using a geographic information system (GIS) and computerized data analysis of service performance trends over time and geography, builds an expectation of learning, improvement, and accountability from agency managers.

Within a few months of O’Malley’s inauguration, CitiStat was instituted across city agencies. It was soon supplemented by a 311 call center system, which allows citizens to call in with complaints or suggestions, linked to a citywide database that monitors service demand and tracks the quality and timeliness of response.

In our view, this is a model system for citizen input and government accountability (although citizen input came only in response to existing service rather than by defining what was demanded in advance). While all the cities we visited had some success in instituting meaningful performance measurement systems, no other cities we visited used measurement more systematically, more often, and more exclusively.
for management and accountability than Baltimore. Further, subtler efforts at culture change took time in other cities—years—where tradition and culture eschewed efforts that sanctioned nonperformers either directly or indirectly. CitiStat was up and running in a matter of months and had the virtue of also developing managerial talent through the process.

Baltimore city officials claim to have experienced 40 percent turnover among top management in agencies, a result that would have been unacceptable in a culture such as Phoenix or Portland. In this system, CitiStat director Mike Gallagher boasted, "managers comply or they're out." Despite some initial instability as departments changed hands, performance improved because those who remained and were promoted were committed to the method of organizational learning and accountability CitiStat allowed. Baltimore was helped in this effort by a strong mayor form of government, capable of ensuring this level of accountability, a form that no other study city enjoyed.

Since 2005, San Diego has changed to a strong mayor form, elected Republican mayor Jerry Sanders, and in 2007 began instituting a top-down performance measurement system for all city agencies. San Diego's early efforts with performance measurement and use of citizen focus groups to help in their development came to an abrupt end during the fiscal crisis of 2004. Efforts have only recently been renewed. Nothing survived from earlier efforts, and there are no plans to engage citizens in their current development. Fiscal and political crises can wreak havoc on developing measurement and management efforts, dependent as they are on the political will and commitment of a new leader.

Worcester and Des Moines notwithstanding, citizen engagement is more often an afterthought in cities' performance measurement efforts than central to them. While a recent study indicated that 70 percent of the largest cities report their performance to citizens, engagement of citizens more directly in identifying the performance that matters to them is far less common (Ho and Ya Ni 2005). A recent report published by the GASB on effective communication of performance measurement revealed through citizen focus groups that citizens are interested in performance information that is important to them and reflects their involvement in the development of the measures themselves (Fountain 1991; GASB 2003; Ho and Coates 2004). Few jurisdictions engage citizens at this level. In fact, as we have shown, citizens are lucky if the results are communicated to them at all (GASB 2002a).

Even reform-minded cities with a history of citizen trust in government were not particularly likely to engage citizens in the performance measurement effort or to be pushed by organized citizen interests to do so. Portland clearly has a strong set of civic and community institutions, a progressive tradition and highly competent public administrators, but formal citizen engagement in performance measurement performance is limited to annual citizen satisfaction surveys and more passive efforts to make annual performance (SEA) reports available.

**Bringing Citizens Back In**

Worcester and Des Moines both sought at the outset to engage citizens as a primary goal of their performance measurement efforts. Both were initiated with funding from the Sloan Foundation, a major sponsor of citizen-initiated demand efforts. Both represent innovative attempts of similar sorts to define citizen engagement as an effort that begins with citizens identifying what service outcomes interest them and contributing to the measurement development process. These efforts are at one end of the continuum of participation, while Portland's would be farther toward the other end, and Baltimore's is a bit of a hybrid.

These are unusual because the initiative for engaging citizens and promoting performance measurement relies on critical partnerships among multiple constituencies. They are designed to "shift the power of information back to citizens by engaging them in the design and use of performance measurement, so that performance measures reflect their concerns and priorities and can be used by policymakers in decision-making" (Ho and Coates 2004, 31). Both efforts had similar goals and strategies, but each found itself in a different cultural and political environment, allowing them to exploit different opportunities and local capacities.

Worcester's effort was initiated by a respected civic organization, the Worcester Regional Research Bureau, a municipal research group funded and supported largely by the city's business community. Independent and nonpartisan with a 20-year history of research on government accountability, effectiveness, and efficiency, the bureau began to examine the performance measurement efforts of other cities to see whether any could aid its own efforts to assess city performance in Worcester. In 2001, the city manager had begun to develop a performance-based budget but found it difficult to engage the city council in its use. The challenge for the city was to agree on the relevant measures in each of the areas; the approach was to structure a series of citizen task forces, with public officials, neighborhood associations, community development corporations, as well as community leaders from the nonprofit and private sector, to suggest meaningful measures that reflected valuable outcomes.
The bureau has recently published the city performance data on each of the developed measures in a set of five reports. Over time, these measures will serve to assess the city and agency progress on key public outcomes. But because of budgetary contraction and staff constraints in the city's Office of Neighborhood Services, a citizen satisfaction survey was undertaken by the bureau as well, because it was viewed as being crucial to the performance measurement effort and would otherwise have been discontinued. An adjacent to this effort, (which has been well received as its first reports were released) is ComNET, a neighborhood-based project in which citizens engaged in an effort to define their local neighborhood service priorities and selected indicators of service outcomes (neighborhood physical conditions) that are monitored in several pilot neighborhoods around the city. As in Des Moines, local citizens have been trained to use handheld computers and digital cameras to record and report ratings on neighborhood conditions—such as street conditions and abandoned cars—and assess the city and agency progress on key public outcomes. Data on performance from ComNET are electronically reported to appropriate city agencies for remediation after review by citizen groups. In 2007, 12 neighborhoods surveyed more than 180 miles of the city, where about a third of its population lives. Agencies are now working to use the data to help allocate maintenance and repair resources and realign priorities.

While the city is engaged and has been supportive of the effort, the bureau is clearly the driver—and, in fact, as a watchdog, it became the “scorekeeper,” (a term the bureau has used itself). But the bureau’s hope is that creating the data will help engage city agencies and politicians to use it in crafting policy. It also hopes it will allow more citizens and organizations to hold the city accountable.

The impact of the performance measures is significant. All major constituents reported this in our interviews. Local community participants reported that the city has used the measures to allocate resources and realign priorities. Since the project launched, they said, response rates have gone up. A culture of measurement is slowly changing the operations of city agencies. Indeed, a recent bureau report proposed the development of a CitiStat management system in Worcester. Even so, community leaders report that they struggle to get teams of citizens trained and willing to walk the relevant routes each year, a factor that could compromise future efforts.

Worcester community organizers are not alone. Even in a city like Des Moines, where the neighborhood movement is very strong, engaging citizens in performance measurement efforts was difficult and required considerable help and commitment (resources and expertise) from the city, the university, the Fund for the City of New York, and the Sloan Foundation. Des Moines summarizes its data in an annual report, focusing largely on workload (not outputs or outcomes) and benchmarks against comparable cities. The report also includes annual citizen satisfaction surveys, which it links in the report to its performance data.

**Obstacles to Performance Measurement**

Exemplary performance measurement systems are costly to develop. Although incipient performance measurement systems can be developed using regularly available data, even cities applauded for exemplary efforts identify these costs as obstacles (GASB 2002a). New systems require significant managerial investments in measurement design and staff training for collection, use, and reporting. In order to stay relevant, measures and reports must be continually refined and altered in response to changing goals and lessons. And multiple constituencies (including citizens!) need to be trained in their use and value for purposes of ensuring that planning, operations, and budgeting rely on analysis of performance data. In all our cases, except Baltimore, city managers cited these as significant costs that often impeded implementation and ongoing improvement, especially during budget downturns.

But traditional cultural norms operating in public bureaucracies may form a more fundamental and challenging obstacle. Most public administrators and agency heads we interviewed acknowledged the discomfort and threat experienced by many old line managers accustomed to managing their operations in conventional ways. Unsurprisingly, measures are generally not used much to manage or budget. Accountability to new sets of performance outcomes and the possibility of sanctions provide implicit motives for obstinacy or subversion. An effective introduction of performance measurement and performance-based management requires a culture change and takes a committed leader with considerable skill willing to provide significant managerial investment and rewards. Of the jurisdictions we visited, only Baltimore actively used negative sanctions to hold managers accountable.

A recent interview with an official in the auditor’s department in Phoenix and elsewhere reinforced the fears associated with the active use of performance data for accountability and convinced us that even mature and well-designed systems implemented in progressive and culturally hospitable environments can be ignored, if not subverted, by anxious city bureaucrats. In newly developed systems such as Baltimore’s, the election of a new mayor without a commitment to measurement could easily undo years of development (and as we will see in our next section, Mayor O’Malley’s replacement, Mayor Sheila O’Malley’s...
Dixon, appears less committed to CitiStat than her predecessor. Even where lip service is paid to existing measurement systems, it is easy to subvert or neutralize the value of the information by agencies and executives not fully committed to their revelations, especially if responsible managers believe they do not have control over the results. The temptation to suppress or manipulate negative data may be particularly strong in situations of significant downturn or when the integrity of the data is suspect.

The development, maintenance, and use of performance measurement that is meaningful, comprehensive, and used for the purposes for which it was designed appears to be less frequent than surveys of local governments would suggest. Political, economic, and cultural factors affect the degree of support, the level of investment, and the ongoing commitment to its refinement and use among various constituencies. Powerful champions inside city government are needed. And considerable evidence exists from our cases that changing political and budget environments exert powerful influence on the ebb and flow of success.

Performance measurement can have champions from many quarters. And while internal governmental support is clearly necessary, efforts in Worcester and Des Moines to encourage external players to take leadership and/or collaborate with city management and elected officials can represent another and perhaps more important means to promote and support performance measurement, especially when the principal goal is accountability to citizens. Citizen engagement, however, takes significant investment and must be nurtured; it is dependent on high levels of trust. While citizen-initiated demand is potentially most likely to result in accountable government, citizen-dependent efforts can be difficult to mobilize and sustain. Our study suggests that citizen efforts can be episodic, driven by a particular service need or crisis, and lacking in long-term potency.

Performance Management

If the promise of performance measurement to improve state and local performance has fallen a bit short in helping to focus attention on results citizens want, it has also delivered less than expected on its promise to improve management. Performance measurement has many functions, but accountability to citizens and managing for results are two prized outcomes that have been expected from its expansion over the last decade and tie directly to the commission’s aspirations. While there has been a lot of activity, the impact of performance measurement has been more difficult to demonstrate. As we have seen, managing for performance or results demands more than just having the capacity to do so. As many have argued, moving from measurement to management requires the will, the power, and the exercise of leadership to achieve at both the state and local levels (Behn 2004; Ingraham, Joyce, and Donahue 2003; King, Zeckhauser, and Kim 2004; Moynihan 2008; Sanger 2008).

The performance management movement (some say “revolution”) is a significant trend in governments across the world (Abramson, Breul, and Kamensky 2006; Kettl 2000). The change driving current government reforms, at its core, turns the conventional bureaucratic paradigm on its head. This postbureaucratic paradigm has evolved in response to changing needs and expectations for governance that the Winter Commission anticipated (Barzelay 1992). “In generic terms, the shift [that Barzelay describes] is from public interest to results that citizens value; from efficiency to quality and value; from a justify-costs mentality to a deliver-value mentality; from functions, authority, and structure to identifying a mission, services, customers, and outcomes; and the like. This paradigm shift suggests an increased interest in system learning” (Battalino, Beutler, and Shani 1996).

Performance measurement has many functions, but accountability to citizens and managing for results are two prized outcomes that have been expected from its expansion over the last decade and tie directly to the commission’s aspirations.

“...highly bureaucratic systems have overly defined and constrained role sets for individuals, and as a result, individuals and teams are limited in their ability to learn, adapt and change.”

Results-oriented leadership seeks to shape the culture of governments or organizations to focus on results and value, mobilize initiative, encourage learning, and promote experimentation to achieve performance. As the commission’s recommendations 15 years ago recognized, these kinds of organizations need different structures and cultures: They need leadership with the authority to empower employees by reducing rules and increasing discretion, rewarding innovation, and tolerating well-conceived failures.

New values must be communicated and championed by management and embraced deep within the organization or jurisdiction in order to change behaviors and ways of working. But...
individuals and teams are limited in their ability to learn, adapt and change” (Barzelay 1992; Battalino, Beutler, and Shani 1996; Osborne and Gaebler 1993). An essential dilemma for leadership in transforming jurisdictions or agencies exists: how to structure the system and exercise leadership to modify behavior and allow contributors to learn and adapt to new values when they have been acculturated and reinforced for staying out of trouble and doing their jobs within their narrow band.

Over the last decade and a half, the movement has resulted in dozens of state and local governments and a host of public agencies undertaking heroic efforts to “reinvent” themselves in a postbureaucratic way. Leadership and its capacity to change organizational culture appear central to all their success in establishing enduring change (Barzelay 1992; Kotter 1995; Osborne and Gaebler 1993; Rainey 2003). But changing an organizational culture is a significant challenge; often efforts are still foiled by the recalcitrance of deeply seeded bureaucratic norms and inhospitable authorizing environments. Interviews with public managers in our six-city sample overwhelmingly cited the difficulties in persuading managers to use performance data in their management strategies. The political resistance inside and outside city agencies and the serious constraints of rigid and centralized budgeting and civil service systems identified as obstacles by the commission 15 years ago are still apparent in some quarters.

**Getting Elephants to Dance**

We have likened the challenges of transforming public bureaucracies to nimble and innovative performance-based systems to getting elephants to dance (Sanger 2007). Optimism, however, is warranted. Increasing numbers of jurisdictions and state and local agencies have launched performance-based management efforts around the country. Some efforts began as early as the commission’s recommendations. Since then, there has been a proliferation of efforts seeking to manage for results. Because of the sheer number and variety of institutions adopting these approaches, we examined several state and agency efforts to supplement our analysis of the six city case studies.

We consider one state effort (Iowa), one citywide effort already discussed (Baltimore), and a number of city agency efforts—two in New York City (Department of Finance and the Police Department) and one in Philadelphia (school district). These examples are illustrative of the types of efforts being launched around the country; unfortunately, no comprehensive survey of performance management efforts in state and local government have yet been undertaken, nor have there ever been systematic efforts to evaluate the relative success of various approaches. Further, as we have seen, the investment and endurance of these efforts are often tied to leadership, but the survival of the strategies is rarely charted.

Inevitably, therefore, our enthusiasm for the increasing investment in performance management systems, which appear to respond so directly to the challenges of contemporary governance (Kettl 2007), will require more rigorous judgments based on a systematic analysis of large samples. We have not attempted that here. This section describes the characteristics associated with a variety of performance strategies which have or promise significant performance improvements. We reflect on their value and their promise as efforts proliferate further.

Iowa represents an early initiator of reform that began in 1993. Through an evolutionary process that involved the leadership of three governors, the state has embraced a performance management framework with multiple elements: benchmarks, results-oriented program performance measures, an investment budget model, and budgeting for results that ties budget allocations to desired results. Iowa’s Accountable Government Act of 2001, passed under Governor Tom Vilsack, established five measurable goals to which state agencies’ performance is linked, embraced results-based budgeting, and offered agencies that were willing to employ the measures greater budget and operating flexibility. This element won an Innovations in American Government Award in 2005. Leadership derives from the governor’s office, and measurement and accountability for results are reinforced by quarterly reviews. Annual agency performance reports are public. Results have been significant in terms of budget savings and improved performance, but they vary by agency.

The New York City Department of Finance, under a charismatic and bold commissioner, embraced another performance management strategy known as a balanced scorecard and dubbed it “SMART” (Kaplan and Norton 1997). Adapted from use in the private and nonprofit sectors, the balanced scorecard is a performance measurement and strategic management system that translates an organization’s mission and strategy into a balanced set of integrated performance measures. The scorecard tracks costs, results, customer satisfaction, and employee engagement, among other measures. The implementation process engages the entire agency in a collective effort: to translate the vision and gain consensus, communicate objectives, set goals, link strategies, set performance targets, allocate resources, and establish milestones. In addition, the design provides and institutionalizes opportunities for continual feedback and learning (Chan 2004).

The rest of our examples, including Baltimore described earlier, represent variants of a class of performance strategies characterized as PerformanceStat
Most successful efforts studied at all levels of government identify the critical challenges to success as on the “demand side.” Legislation and administrative mandates can require and create processes—the requirements for strategic plans, reporting systems, creating performance measures—but, as we have seen, ensuring the use and value of these processes is less certain. Any effective use of these for management requires strong and committed leadership, both from the top down and the bottom up (Behn 2004; Kamensky 2006; Perrin 2006; Sanger 2008). All our cases represent efforts initiated by strong leaders who built ground level leadership deep within their organizations: some through inclusion, training, and rewards (Iowa and the New York City Finance Department), others through ensuring the recognition of the personal and organizational returns from embracing new roles and values (CompStat and CitiStat). When the New York City commissioner of finance adopted the balanced scorecard performance management strategy, she realized early on that a top-down effort to change the culture from doing one’s job well to producing a result would not succeed without other inducements (in fact, she experienced early significant pushback!). For the commissioner and her leadership team, it was a battle for the hearts and minds of the workforce—managerial and line workers alike—and the challenge was to get them to embrace the use of data to learn and experiment with new strategies to achieve concrete results. Communication and explicit training efforts to build worker participation and engagement are critical in the movement to such a system (Beer, Eisenstat, and Spector 1990; Buckingham and Coffman 1999; Duck 1993; Larkin and Larkin 1996; Strebel 1996). In all our cases, the particular means of inspiring staff to change their practices varied, but all used incentives, sanctions, feedback, coaching, and learning vehicles to build capacity and leadership deep in the organization.

Three other elements characterized every successful performance management system in our cases and are recognized in the literature more generally. At a fundamental level, management’s strategic goals ought to drive the measures, and the data must accurately and reliably capture them (Behn 2003, 2004; Metzenbaum 2006). In the New York Police Department’s CompStat system, early knowledge that the incidence of illegal guns was directly correlated with crimes of various kinds linked the reduction of illegal guns to safety and crime reduction, a principal goal of the mayor. Performance measures of illegal guns seized, therefore, became a valuable and useful metric that linked to a most important result for the New York Police Department (Smith and Bratton 2001). The city’s Department of Finance has developed metrics for its agency scorecard that often derive from the relevant measures needed to manage within the agency. Thus, if “helping taxpayers pay the right amount on time” is one of the five agency goals or results on the agency scorecard, then metrics chosen must reflect that result.

Further, to be effective, measures must be clear, accurate, and credible to internal and external audiences. In Baltimore, New York, and Iowa, the challenge of producing reliable data prompted continual adaptations and refinements as the system developed. Systematic feedback, learning, and consulting with a range of stakeholders within and external to the organizations and agencies proved critical. In Iowa, for example, the evolution took more than a decade. Successful performance management efforts do and should evolve and adapt over time to meet the needs of the various constituents and stakeholders who shape the aspirations for particular outcomes and results and, perhaps, new capacities for measurement (Sanger 2007).

Transparency and public reporting of performance were also key elements of all of our successful cases. The sharing of data creates a taste for and expectation of measurement and accountability, which ratchets up pressure for it from outside. It also produces visibility for the governmental or organizational unit and builds internal pride and incentives for continual improvement. In all of these efforts, however, transparency created risks. In exposing agencies or units within them to external scrutiny, transparency may inadvertently create incentives to game, distort or subvert the numbers. It can also make managerial failures more evident. Auditing data and protecting their integrity is crucial.

Public reporting also builds trust and goodwill from authorizers, citizens, customers, and other stakeholders. For example, Philadelphia’s SchoolStat has built visibility for school reform efforts, and the public reporting of improvements has created rising expectations while supporting increased investment in the change effort (Patusky, Shelley, and Botwinik 2007).

Finally, the ultimate value of a performance managed system is its potential to provide feedback and learning.
about operations and the strategies that produce improvement. When metrics are collected in a disaggregated way, like CompStat’s precinct level, or the Finance Department’s borough offices, managers can search for high performers—what Behn (2007) calls “positive deviants.” This allows hypotheses to be developed and tested about what is driving improved performance. Innovation and accountability are two sides of the same coin. All successful performance management systems provide mechanisms for review and feedback that can promote learning. Cultivating initiative and willingness to take well-conceived risk depends on an ability to value their lessons, even when they are not successful.

**Conclusion**

The Winter Commission’s aspirations for improvement in American governance at the state and local levels recognized the increasing inadequacy of the institutions and systems built in earlier eras to meet contemporary challenges. Modernizing these public institutions and systems, in part, requires us to embrace a new paradigm based on a new set of values. The ability to identify what we want from government, to measure the results we seek, and to shape policy accordingly has been at the center of the performance measurement and management quest in recent years.

The good news is that much has been learned, and measurement systems that can report results to citizens are growing. The development of voluntary standards for performance reporting by the GASB will generate norms that create a taste for and expectation of measurement and accountability; it ratchets up the pressure from the outside to report on results and to perform. Further, when citizen engagement in defining the results that matter becomes the standard for performance reporting, increased accountability will be achieved. This is happening already. But as we have seen, the capacity to manage for performance, even with better measures, does not ensure its use. Greater visibility and understanding of successful performance management strategies does create demand. CitiStat is the next generation of CompStat. Four principle lessons have emerged from our research:

- Performance measurement is growing in states and local governments, but more often without the engagement of citizens and with unrealized use for management.
- Some improvement is evident at all levels of government, but cities do better than states and performance management efforts are growing most successfully at the agency level.
- Jurisdictions and agencies with the best performance reporting and performance management efforts have strong mission driven leaders at the helm who communicate the mission, motivate employees, shape strategies, and provide support, rewards, and sanctions for achievement.
- The state of knowledge about what jurisdictions and agencies are doing, why, and with what success is growing but remains inadequate to inform intervention or policy.

Scholarship on what works and why can help promote the institutions and systems for twenty-first-century performance management. But we do not know enough about the quality and value of performance measurement in state and local government and its relationship to outcomes citizens want. Research on the determinants of successful implementation, adoption and use of performance measurement for management is in its infancy. Even when performance measurement is used, the impact of performance management systems on outcomes is generally unmeasured and little is known about its cost-effectiveness or endurance over time. It is not well understood how these efforts develop and survive over time, and small samples make generalization difficult. Contextual variables matter, but too little is known about the relative importance of context, processes, institutional arrangements, and leadership in explaining what drives performance and its institutional embeddedness.

Finally, scholars may find more fruitful avenues of exploration in the use of performance measurement and management at the agency level, where considerable success and innovation can be found. Promotion of best practice may indeed have more power at the agency level, where professional management and less political conflict over goals may be found. Scholars may be looking for success in the wrong places. The promise is there. But as many analysts have warned, our expectations should be tempered. Recognizing the limits as well as the potential of the performance movement is also important.

**Notes**

1. Multiple organizations are committed to improving public management and public accountability, such as the International City/County Management Association (ICMA), the Association of Government Accountants (AGA), the Government Finance Officers Association (GFOA), the American Society for Public Administration (ASPA), and the Government Accounting Standards Board (GASB), as well as
foundations and research organizations such as the Alfred P. Sloan Foundation, the Fund for the City of New York, and the Urban Institute. They all have been involved in promoting the use and improvement of performance measurement through research and documentation of efforts, training, technical assistance, grants, awards, and collaborative activities such as benchmarking efforts and other “promotion” activities. Some specific initiatives include the AGA’s Certificate in Excellence in SEA Reporting Program; the Fund for the City of New York’s continuing development and implementation of methodologies that provide reliable, nonpartisan, objective, citizen-based assessments of the effectiveness of New York City government agencies; Iowa State University’s study of city-elected officials’ attitudes toward performance measurement; and the National Civic League’s effort to increase citizen participation in performance assessment in jurisdictions involved in the ICMA’s Local Government Comparative Performance Measurement Project.

2. For example, the ICMA, GASB and ASPA.

3. State institutions are complex; their array of services is broad, but their delivery indirect. Politicians and managers are less closely connected with the citizens they serve and the frontline employees who ultimately deliver them than are city governments. For state politicians, performance data do not make political decisions simpler but simply add an additional level of information to incorporate (see Moynihan 2008).

4. This is not entirely true because, as Behn (2003) reminds us, we need different measures for different purposes. But performance management systems rely on clarity of mission, and goals and outcome metrics most often clearly communicate what the agency is trying to accomplish. Output measures might not be preferable for use when managing workers deep in the organization, where output measures that “roll up” to outcomes might be more helpful.

5. We contacted organizations such as the GASB and ICMA, the Sloan Foundation, the Urban Institute Neighborhood Network Indicators Project, the Maxwell School’s Government Performance Project, and the Fund for the City of New York, as well as city Web sites.

6. Baltimore also had the unique virtue of having a parallel nonprofit-funded and managed system that was developing comprehensive community indicators at the neighborhood level and that city agencies contributed to and cooperated with. The Baltimore Neighborhood Indicators Alliance uses neighborhood based indicators data to help organize and promote well-informed decision making for community change.

7. Phoenix produces several reports, such as the City Manager’s Executive Report, the City of Phoenix Community Attitude Survey, and the Performance Achievement Program report, which all publish performance measurement results.

8. Phoenix reports police data on a submetropolitan basis, and Portland publishes its satisfaction survey on a neighborhood basis, but neither reports performance by neighborhood.

9. Portland’s commission form of government has five elected commissioners who serve as both the legislative body and as administrators of the various operating departments, which the mayor (who is one of them) assigns to them. The decentralized nature of Portland’s government by definition makes the performance measurement process a collaborative one. The independently elected auditor has played a key role and is the operational force in the development and reporting of performance outcomes.

10. CitiStat also uses a computerized 311 call center for all service requests and complaints, which are immediately transferred to the relevant department for follow-up with specific response times associated with each request. These, too, are geographically identified. CitiStat staff survey 100 callers each week to follow up on service responsiveness and satisfaction, and these data are also used to evaluate departmental performance.

11. The original motivation for our research funded by the Sloan Foundation was to identify the factors related to citizen engagement in places where it was evident in order to help promote and support performance measurement efforts that were citizen-centric.

12. Des Moines was one city among nine in Iowa that participated in Citizen Initiated Performance Assessment, in which citizens, elected officials, and departmental staff were involved in a systematic way to develop and select performance measures for city services (see Ho and Coates 2004).

13. Baltimore has a separate but related social indicators project that engages citizens in neighborhoods around the city in strategic planning efforts that are data driven. The Baltimore Neighborhood Indicators Alliance is a cooperative effort of diverse public, private, and nonprofit groups to promote, support, and make better community decisions. It makes available diverse sets of indicators and data on Baltimore neighborhoods (including CitiStat performance data) that help citizens understand their current status and conditions for use in strategic planning, organizing, and monitoring progress in important areas.
of neighborhood interest. We do not evaluate this effort specifically here, but it is clearly complementary and supportive of citizen participation.

14. Fountain (1991) points out that most places have the relevant data needed, and reporting the results would only require a marginal amount of work.

15. Data derive from a variety of sources including public documents and agency and jurisdiction Web postings, academic and professional literature, site visits and case studies (in the case of the New York Department of Finance and Baltimore), and interviews with participants and observers.

16. The Innovations in American Government Award is given annually by the Ash Institute at the John F. Kennedy School of Government at Harvard to exemplary efforts that are identified as best practices and innovative leadership models.

17. CompStat was the first in a group of performance management efforts replicated in various cities and agencies around the country. Behn (2007) has identified two states, eight cities, three counties, and numerous police departments and other city and federal agencies that have adopted variants of the PerformanceStat strategy.

18. Iowa began its performance management efforts in 1993 with a legislative bipartisan effort that established the Council on Human Investment to identify the goals and character of a performance management system, budgeting for results and benchmarks. Three governors later, Iowa has built a state-of-the-art performance management system, an element of which won the 2005 Innovations in American Government Award.

References


——. 2004. Use and the Effects of Using performance Measures for Budgeting, Management, and


2009 ASPA Annual Conference

ASPA's 2009 Annual Conference will be held March 20-24 in Miami, Florida at the Hyatt Regency Miami. The conference will explore and develop innovative practices for public administration to address the growing diversity of communities around the country.

It emphasizes the importance of context and culture on the sustainability of innovations, and ASPA's role in supporting public administration globally from a U.S. perspective. More than 400 proposal submissions were received for the conference! The deadline for submitting abstracts has closed, but for more information on the conference go to:

http://www.aspanet.org/scriptcontent/index_aspaconference.cfm