Have Cities Shifted to Outcome-Oriented Performance Reporting?—A Content Analysis of City Budgets

ALFRED TAT-KEI HO and ANNA YA NI

"Outcome" orientation has been emphasized in performance measurement and reporting in recent years. Using budgetary documents of the largest cities in the United States, this paper analyzes more than 4,800 performance measures reported by 21 cities to show that clear progress toward outcome-oriented performance measurement has been made. It also shows that the selection of performance measures differs among types of municipal services and is driven partly by professionalism in city management and the influence of professional organizations. The paper concludes by discussing the implications of the findings and potential future improvements of performance reporting.

INTRODUCTION

For many decades, performance measurement has been used as an internal informational tool to evaluate departmental operations and make program and budgetary decisions. Its use has primarily been focused on output, workload, and cost-efficiency, and the targeted users have been largely internal staff, such as managers and departmental staff.[1] However, the "reinventing government campaign" in the late 1980s, various federal legislative initiatives, and the educational efforts of various professional organizations have

Alfred Ho is an associate professor at the School of Public and Environmental Affairs, Indiana University-Purdue University Indianapolis. His research focuses primarily on performance measurement, state and local finance and policies, and e-government development. He can be reached at altho@iupui.edu.

Anna Ni is a Ph.D. student and a research associate at the Maxwell School of Citizenship and Public Affairs, Syracuse University. Her research areas include organization theories and management, information technology management, and e-government development. She can be reached at yani@mailbox.syr.edu.

1. Richard J. Fischer, "An Overview of Performance Measurement," *Public Management* 76, no. 9 (1994): 2–8; Thomas P. Lauth, "Performance Evaluation in the Georgia Budgetary Process," *Public Budgeting & Finance* 5 (1985): 67–82; Clarence E. Ridley and Herbert A. Simon, *Measuring Municipal Activities: A Survey of Suggested Criteria and Reporting Forms for Appraising Administration* (Chicago: The International City Managers' Association, 1938).

changed the direction of performance measurement in recent years.[2] Government officials are now encouraged to examine "outcomes" and adopt "results-oriented" performance measurement.[3] They are also challenged to expand the readership of performance measures and budgetary information and involve more stakeholders, such as the community leadership and the general public.[4]

While many researchers have examined why the shift is important and necessary, not many empirical analyses have been done to show the extent to which it has occurred. The purpose of this paper is to analyze how some of the largest U.S. cities report performance measurement results. Specifically, we are interested in the following questions:

- What types of measures are reported?
- How do cities integrate performance measures with other information, such as explanatory information and program goals?
- Is there any systematic pattern in performance reporting among cities? For example, do the types and quantity of measures vary by jurisdiction type, size, or region?

In the following, we first examine why many professional organizations have recommended outcome measurement in recent years. We then explain the empirical methodology and the classification system used here to analyze performance measures. Based on our empirical results, we discuss the implications for future development of performance budgeting and reporting.

RECENT SHIFT IN THE RECOMMENDED PRACTICES OF PERFORMANCE REPORTING

As early as the turn of the twentieth century, the New York Bureau of Municipal Research had already recommended the use of performance measurement to evaluate the

2. John Mikesell, *Fiscal Administration*, 6th ed. (Belmont, CA: Wadsworth/ Thomson Learning, 2003).

3. Jens K. Kristensen, Walter S. Groszyk, and Bernd Bühler, "Outcome-Focused Management and Budgeting," *OECD Journal on Budgeting* 1, no. 4 (2002): 7–34; Peter Smith, ed., *Measuring Outcome in the Public Sector* (London: Taylor & Francis, 1996); Fred Thompson, "Mission-Driven, Results-Oriented Budgeting: Fiscal Administration and the New Public Management," *Public Budgeting & Finance* 15, no. 3 (1994): 90–105; U.S. General Accounting Office, *Results-Oriented Budget Practices in Federal Agencies*, GAO-01-1084SP (Washington, DC: U.S. GAO, 2001); Joseph S. Wholey and Kathryn E. Newcomer, "Clarifying Goals, Reporting Results," *Using Performance Measurement to Improve Public and Nonprofit Programs*, ed. Kathryn E. Newcomer (San Francisco: Jossey-Bass, 1997): 91–98.

4. Alfred Tat-Kei Ho and Paul Coates, "Citizen Participation: Legitimizing Performance Measurement as a Decision Tool," *Government Finance Review* 18, no. 2 (2002): 8–10; Andrew F. Long and Jenny Jefferson, "The Significance of Outcomes within European Health Sector Reforms: Towards the Development of an Outcome Culture," *International Journal of Public Administration* 22, no. 3 & 4 (1999): 385–424; Irene Rubin, "Budgeting for Accountability: Municipal Budgeting for the 1990s," *Public Budgeting & Finance* 16, no. 2 (1996): 112–132.

"economy" and "efficiency" of government operations.[5] For many decades, this was the focus of performance measurement, which was designed and used by managers to analyze the output–input relationship and find means to enhance the cost-efficiency of service delivery. However, the economic orientation of performance measurement was soon found to be insufficient for a number of reasons. First, budgetary decision-making is a political process.[6] Although objective economic analyses, such as unit-cost analysis and efficiency analysis, are helpful to enable managers to understand how to best allocate resources, decision-makers often need to balance political and economic rationales. What services the government should provide, how these services should be provided, and how stakeholders perceive the priorities of service delivery and the quality of services are political questions that cannot be easily captured by pure economic analysis.

The attempt to expand performance measurement beyond pure economic analysis was reinforced by the "planning–programming–budgeting system" (PPBS) reform of the 1960s. PPBS shifted the focus of budgeting from control and cost-accounting to planning and program management. As budgeters and managers asked the questions, "What are the goals of the programs?" "What do we want to accomplish through these programs?" and "Who benefits and who loses from this program?" they inevitably needed to focus on measuring other aspects of services, such as output, effectiveness, and outcomes.

In the 1970s, the Urban Institute and the International City Management Association (ICMA) produced a series of publications to help managers think about performance measurement in terms of "effectiveness" and results.[7] They also began to encourage greater utilization of citizen surveys and other feedback mechanisms in performance measurement. In the 1980s, calls for outcome measurement became even more strident as a result of the "reinventing government" movement. In response to the rising cynicism about public agencies, government reformers challenged policymakers and managers to think beyond bureaucratic boundaries and become more mission- and customer-driven and more results-oriented. For example, Osborne and Gaebler have cited numerous examples from local governments to show the importance of measuring citizen satisfaction and results accomplishment, and of budgeting "entrepreneurially" by output and outcomes.[8] The federal government also contributed indirectly to the momentum of this reform by passing several pieces of landmark legislation, including the Chief Financial Officers Act of 1990 and the Government Performance and Results Act of 1993, to

5. New York Bureau of Municipal Research, "The Citizen and the Government. A Statement of Policy and Method," *Municipal Research* 57 (1915): 1–4.

6. Aaron Wildavsky and Naomi Caiden, *The New Politics of the Budgetary Process*, (New York: Longman, 2004); Irene Rubin, *The Politics of Public Budgeting: Getting and Spending, Borrowing and Balancing* (New York: Chatham House, 2000).

7. Harry Hatry, et al., *How Effective Are Your Community Services? Procedures for Monitoring the Effectiveness of Municipal Services* (Washington, DC: The Urban Institute, 1977).

8. David Osborne and Ted Gaebler, *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector* (New York: Penguin Books, 1992).

strengthen the link between goal-setting, program management, and performance measurement in federal agencies.[9]

It was in this context that many professional organizations advocated outcome measurement in the 1990s. For example, the Governmental Accounting Standards Board (GASB) issued Concepts Statement No. 2 on "Service Efforts and Accomplishments Reporting" in 1994, in which it explained the importance of outcome measurement to enhance the public accountability of governmental agencies.[10] The Urban Institute, which has continued to be an active supporter of this movement throughout the past two decades, promoted the usage of "outcome-focused performance measurement systems."[11] The National Advisory Council on State and Local Budgeting (NACSLB), which consists of several major national professional organizations related to public budgeting and financial management,[12] also encouraged the shift toward "outcome-oriented" measurement. In its 1998 report, the NACSLB stated,

> A good budget process incorporates a long-term perspective, establishes linkages to broad organizational goals, focuses budget decisions on **results and outcomes** [emphasis added], . . . [T]he budget process is not simply an exercise in balancing revenues and expenditures one year at a time, but is strategic in nature, encompassing a multi-year financial and operating plan that allocates resources on the basis of identified goals.[13]

In addition, advocates of performance measurement challenged managers and budgeters to involve more stakeholders, including citizens, community leaders, and elected officials, in the development and usage of performance measurement, so that the "outcomes" being measured would be more relevant and meaningful to decision-makers.[14] Bouckaert summarizes the changes of the performance measurement campaign in the 1990s as follows,

---

9. Philip G. Joyce, "Using Performance Measures for Federal Budgeting: Proposals and Prospects," *Public Budgeting & Finance* 13, no. 4 (1993): 3–17; U.S. GAO, *Performance Budgeting: State Experiences and Implications for the Federal Government*, GAO/AFMD-93-41 (Washington, DC: GAO, February 17, 1993).

10. Governmental Accounting Standards Board, *Concepts Statement No. 2 on Concepts Related to Service Efforts and Accomplishments Reporting* (Hartford, CT: GASB, 1994).

11. Harry Hatry, *Performance Measurement: Getting Results* (Washington, DC: The Urban Institute, 1999).

12. The National Advisory Council on State and Local Budgeting consists of the Association of School Business Officials International, the Council of State Governments, the Government Finance Officers Association (GFOA), the International City/County Management Association, the National Association of Counties, the National Conference of State Legislators, the National League of Cities, and other business and academic membership.

13. National Advisory Council on State and Local Budgeting. *Recommended Budget Practices. A Framework for Improved State and Local Government Budgeting*. Chicago: National Advisory Council on State and Local Budgeting, 1998.

14. Governmental Accounting Standards Board, *Report on the GASB Citizen Discussion Groups on Performance Reporting* (Norwalk, CT: GASB, July 2002).

> The focus has shifted from how much money was spent according to laws and regulations, to what happened to these resources, how they were used, and with what result for society. Information produced by such performance measurement systems becomes part of a process of management, which is defined as taking responsibility for the performance of a system. It is used as a major element in the mechanism of accountability for administrative or political hierarchies. . . . Its use has passed beyond internal levels to external fields.[15]

METHODOLOGY OF THE STUDY

Previous studies have focused on the rationales favoring the shift in performance measurement objectives and utilization without examining the extent of the trend itself. The purpose of this study is to examine how outcome measurement and reporting has been practiced by U.S. cities. Most empirical studies that examine the practice of performance measurement rely on self-reporting surveys.[16] While this method is more convenient and efficient, it relies totally on the respondents' understanding and interpretation of "outcome," "output," and "input" measures, which may create potential measurement errors. Also, most of these studies only ask whether a jurisdiction collects certain types of measures and do not examine the scope of usage of each type of performance measures, such as the number of measures collected and reported and the relative ratios of outcome, output, input, and efficiency measures. Finally, survey methods cannot provide an in-depth examination of how performance measures are formatted for readers and the extent to which they are integrated with other budgetary and policy information, such as program goals, performance benchmarks, and other explanatory information.

---

15. Geert Bouckaert, "Measurement and Meaningful Management," *Public Productivity and Management Review* 17, no. 1 (1993): 321–343.

16. See Evan Berman and Xiaohu Wang, "Performance Measurement in U.S. Counties: Capacity for Reform," *Public Administration Review* 60, no. 5 (2000): 409–420; Alfred Ho, "Perceptions of Performance Measurement and the Practice of Performance Reporting by Small Cities," *State and Local Government Review* 35, no. 3 (2003); Meagan M. Jordan and Merl M. Hackbart, 1999; Julia Melkers and Katherine Willoughby, "Budgeters Views of State Performance Budgeting Systems," *Public Administration Review* 61, no. 1 (2001): 54–64; Theodore Poister and Gregory Streib, "Performance Measurement in Municipal Government: Assessing the State of the Practice," *Public Administration Review* 59, no. 4 (1999): 325–335; Gregory Streib and Theodore Poister, "Performance Measurement in Municipal Governments," *ICMA Municipal Year Book 1998* (Washington, DC: International City/County Management Association, 1998); Xiaohu Wang, "Perception and Reality in Developing an Outcome Performance Measurement System," *International Journal of Public Administration* 25, no. 6 (2002): 805–829; Katherine G. Willoughby and Julia E. Melkers, "Implementing PBB: Conflicting Views of Success," *Public Budgeting & Finance* 20, no. 1 (2000): 105–120, among many other studies. An exception is a study by Donald P. Moynihan and Patricia W. Ingraham, "When Does Performance Information Contribute to Performance Information Use? Putting the Factors in Place," Working paper of the Campbell Public Affairs Institute, Maxwell School of Citizenship and Public Affairs, Syracuse University (2001). However, the content analysis of this study focuses on the "quality" of performance information in state documents, such as clarity of goals and consistency of measures, and does not analyze the types of measures reported and the format of reporting.

This study utilizes a content analysis of the budget documents of the 30 largest U.S. cities to examine how cities report performance measures in eight specific areas—community development and neighborhood services, economic development, fire and emergency medical services, libraries, parks and recreation, planning and zoning, police, and street maintenance and repairs.[17] This methodology allows us to organize and analyze measures by programs and activities and to examine the link between program measures, goals, and budget presentation.

To avoid confusion and measurement error in the analysis, we follow the classification of performance measures suggested by the Urban Institute and the GASB, and define different types of performance measures as follows:[18]

Outcome Measures

Outcome measures capture the results or the consequences of service delivery that are important to the public and customers. There are at least two subcategories of outcome measures:

(a) *End outcomes*. End outcomes measure a program's results—whether they are desired or undesired, intended or unintended. For example, the outcome of police investigation may be crime clearance rates, the outcome of a drug prevention program may be the reduction of drug usage in the target population, and one of the outcomes of an after-school educational program may be the number of youth graduating from high school or enrolled in colleges.

(b) *Intermediate outcomes*. Intermediate outcomes are the intermediate results that are expected to lead to end outcomes. Many managers and policymakers rely on intermediate outcomes to evaluate the performance of a department, because many extraneous factors that are outside the control of government agencies can influence attainment of the end results.[19] An example of intermediate outcomes is the number of voluntary participants completing a training program. Completion is expected to yield the end

17. Because most cities do not separate the budget and performance measures for fire and emergency medical services, the following analysis combines the two services and their performance measures.

18. GASB, 1994; Hatry, 1999.

19. For example, the success of crime reduction by the police department depends significantly on the national economy and the unemployment rate, and the success of educational and training programs for youth is related to the program recipients' family income and educational level. A street repair team may have provided excellent services to the community by fixing the cracks, resurfacing the roads, and removing the snow and ice according to professional standards, however, they cannot guarantee the safety of the roads and prevent all accidents from happening because of the effect of individual drivers' actions. Because of such uncertainty and uncontrollability in achieving the end-outcomes of services, intermediate outcomes are used as proxies to evaluate whether programs are at least accomplishing the necessary steps or elements leading to the end results.

outcome of helping trainees acquire the necessary skills, even though this may not be a guaranteed result. Customer or user satisfaction with the quality of services is another commonly used intermediate outcome measure.[20]

Output Measures

Output measures capture the amount of products and services completed or delivered. Examples of output measures include lane miles of road repaired, the number of programs or facilities provided, and the average number of personal assistance hours per client.

It should be noted that there is often ambiguity between output and intermediate outcome measures in public services.[21] For example, the response time of emergency medical services is a service "output" but can also be treated as an intermediate outcome because it is a significant public concern. The number of arrests by the police department is also an output measure but can be treated as an intermediate outcome measure for law enforcement.

To ensure consistency in the treatment of all performance measures reported in different programs, we adopt the following classification rules in our analysis:

- Customer satisfaction with a program is treated as an intermediate outcome.
- The number of voluntary participants or the number of users of any nonmandatory program who "successfully complete" the program is treated as an intermediate outcome. However, merely reporting the number of participants or customers served is treated as an output measurement since it does not indicate any result or completion of the program.
- Similarly, the number of users of library services is defined as an output measure, even though it can be treated as an intermediate outcome measure.
- The response time of emergency services or replies to customer requests is treated as an intermediate outcome toward the goal of saving lives.
- Measures of service quality, such as accuracy, accessibility, and safety, are treated as intermediate outcomes.
- The number of complaints about service quality is treated as an intermediate outcome.
- The number of arrests and cases cleared or successfully resolved by law enforcement agencies is treated as an intermediate outcome measure.

20. High customer satisfaction is often related to the effectiveness of accomplishing the end goals, even though the relationship is indirect and not guaranteed. For example, students' satisfaction with the instructor of a program should be positively related to students' learning, but high quality in instruction cannot guarantee that all students will pass a qualifying examination, which is an end result.

21. Hatry, 1999.

Efficiency and Productivity Measures

Efficiency and productivity measures assess the ratio of output to input. For a given amount of resources invested, a program is more "productive" or "efficient" if it delivers more outputs to the public. Some common examples of efficiency measures include miles of road paved per dollar spent, average processing cost per case or per client assisted, and average processing time per case or per client assisted.

Many programs also measure the number of clients served per caseworker or the total caseload handled per worker. Since these measures deal with the relationship between an output (number of cases or clients) and the human resource input, they can be regarded as an efficiency measure. However, they can also be interpreted as "output measures" with the service input as a qualifying denominator. Since the intent of most of these measures is to show the "workload" aspect rather than the "efficiency" aspect, we categorize these measures as "output" measures in the following analysis.

Performance-Related Information

In addition to the above performance measures, many cities report other indicators that are not "performance" measures but can provide useful information to citizens about the environmental factors that influence the performance of public services. We categorize these nonperformance indicators in the following ways:

1. *Input measures*. Input measures indicate the amount of resources and organizational efforts invested to provide certain services. Examples of input measures include the amount of spending or grant funding in a program, and the amount of man-hours invested in a program.

It should be noted that whether a measure is an "input," "output," or "outcome" sometimes depends on the nature of the program. For example, "the number of newly hired police officers" is an input measure from a service delivery perspective. However, it may be an "output" measure for an administrative program that tries to recruit and train more new police officers. In our analysis, all measures are categorized from the service delivery perspective, with the exceptions of measures in "administrative support programs." Measures in such programs are categorized from a managerial perspective. For example, completion of a budget, the number of trainings conducted, the number of items purchased, and the number of machines maintained are "output" measures of an administrative unit, even though they do not provide any indication about the ultimate performance of service delivery. Employee turnover rate and the success rate of grant applications are categorized as "intermediate outcomes" of administrative support, even though they are about the quality of "input" factors. Also, the percentage of spending variance in a program budget is a measure of "intermediate outcomes" of budgetary control, even though it gives no information about how well a service is provided to the public. We regard it as an "intermediate" step to the endoutcomes of budgetary administration, which include ensuring that there is sufficient funding for delivering all the planned programs.

2. *Demand measures*. Demand measures reflect the potential workload level or the potential number of clients of a program. Although they are not performance measures, a comparison of the demand measures with the actual program output provides valuable information about how well a department is able to meet the needs of the public. Examples of demand measures include the number of lane miles that need repairs and the number of low-income families eligible for an assistance program.

3. *Explanatory factors*. Explanatory factors measure the environmental factors that are likely to affect the performance of a program. Common types of explanatory factors include the size of a jurisdiction; economic factors, such as economic growth and the unemployment rate; and cost or price factors.

RESULTS OF THE ANALYSIS

In April 2002, we requested the 30 largest cities in the U.S. to send us their latest budgetary documents. Some of the documents were obtained online from the cities' websites, and some were obtained by mail. Among the 30 cities selected for this study, 21 included performance measures in their FY2001 or FY2002 budgets (see Table 1). This does not necessarily imply that the remaining nine cities did not practice performance measurement. For example, New York, Philadelphia, and Phoenix have performance measurement reports that are published separately from their budgets. Similarly, the City of Charlotte, North Carolina, has a separate annual report of city performance and a citizen report of city services, which have extensive reporting of performance measurement results. However, since these cities do not integrate performance measures into their budgetary documents, they were excluded in the following analysis.

Performance Measures Reported by Cities by Service Areas

A total of 4,817 measures from eight different municipal service areas were analyzed in this study. On average, each city reported 31 measures per service area (see Table 2). Community and neighborhood development had the largest number of measures reported—with an average of about 61 measures per city. Given the fact that community and neighborhood development includes many services, such as children and youth programs, community centers, and community beautification and development, and usually involves a lot of voluntary participants, it should not be surprising that many cities chose to report their effort and accomplishments in this area.

The second service area that had the largest number of performance measures was police protection, which had an average of about 42 measures per city. One reason why this service area tends to have many measures is because cities must prepare categorized crime statistics and reports on law enforcement efforts for the Federal Bureau of Investigation, regardless of their decision to have performance reporting in their budgets.

TABLE 1
Cities Included in This Study (by population size)

City Name	Population	Budget and financial documents studied	Report performance measures
New York	8,008,000	City of New York Executive Budget FY 2002	
Los Angeles	3,695,000	City of Los Angeles Budget Summary FY 2000–2001	
Chicago	2,896,000	City of Chicago 2002 Budget Recommendations	
Houston	1,954,000	City of Houston FY 2002 Adopted Budget	Yes
Philadelphia	1,518,000	City of Philadelphia The Mayor's Operating Budget in Brief for FY2002 City of Philadelphia Five-Year Financial Plan FY2002–FY2006	
Phoenix	1,321,000	The Phoenix Budget Summary 2001–2002 City of Phoenix Proposed 2001–2002 Annual Budget	Yes
San Diego	1,223,000	City of San Diego Annual Budget FY 2001 Volume 1 Citizen's Budget	Yes
Dallas	1,189,000	City of Dallas Proposed 2001–2002 Budget	Yes
San Antonio	1,145,000	City of San Antonio, Texas Adopted Annual Operating Budget FY 2001–2002	Yes
Detroit	951,000	City of Detroit Fiscal 2001/02 Budget	Yes
San Jose	895,000	City of San Jose 2001–2002 Adopted Operating Budget	Yes
Indianapolis	792,000	City of Indianapolis 2002 Proposed Budget	
San Francisco	777,000	City of San Francisco Mayor's Proposed Budget 2001–2002	Yes
Jacksonville	736,000	City of Jacksonville, Florida, Annual Financial Plan for the FY Ending September 30, 2001	
Columbus	711,000	City of Columbus 2002 Budget	Yes

Austin	657,000	City of Austin FY 2001–2002 Approved Budget	Yes
Baltimore	651,000	City of Baltimore Fiscal 2002 Summary of Adopted Budget	
Memphis	650,000	City of Memphis FY 2002 Operating Budget	Yes
Milwaukee	597,000	City of Milwaukee 2002 Adopted Budget	Yes
Boston	589,000	City of Boston Operating Budget FY 2002	Yes
Washington, DC	572,000	Government of the District of Columbia FY 2002 Proposed Budget and Financial Plan	Yes
Nashville-Davidson	570,000	The Metropolitan Government of Nashville and Davidson County FY 2001–2002 Recommended Operating Budget	Yes
El Paso	564,000	City of El Paso, Texas, FY 2001 Budget	Yes
Seattle	564,000	City of Seattle 2002 Proposed Budget	
Denver	555,000	City of Denver Proposed 2002 Budget	Yes
Charlotte	541,000	City of Charlotte FY2002 & 2003 Final Strategic Operating Plan	
Fort Worth	535,000	FY2000–01 City of Fort Worth Adopted Budget	Yes
Portland	529,000	City of Portland, Oregon, Adopted Budget, Volume One, Bureau Budgets, Programs and Services, FY 2001–2002	Yes
Oklahoma City	506,000	City of Oklahoma City Annual Budget FY 2001–2002	Yes
Tucson	487,000	City of Tucson Adopted Budget FY 2002	Yes

FY, fiscal year.

TABLE 2
Number of Performance Measures Analyzed, by Service Areas

	Total number of measures reported	Number of cities	Average number of measures
Community and neighborhood development	1,092	18	60.7
Economic development	245	17	14.4
Fire and emergency medical services	821	21	39.1
Library	402	18	22.3
Park and recreation	805	20	40.3
Planning and zoning	338	16	21.1
Police	887	21	42.2
Street maintenance and repairs	227	20	11.4
Total	4,817	Average	31.4

Hence, collecting these data and reporting 40–50 crime-statistics measures is usually not a significant additional burden.

The area that had relatively fewer performance measures reported in the FY2001 or FY2002 budgets was street maintenance and repairs, probably because these operations are often driven by professional technical standards or requirements, and many public works directors who receive engineering training are not accustomed to the idea of performance measurement. On average, a city reported about 11 measures in this area, most of which were output measures.

It should be noted that the number of performance measures reported in a program or departmental budget varied significantly among cities. Austin, for example, had a total of 113 measures in the police budget, while Columbus only had four measures. Boston had 41 measures in economic development, while many cities, such as Columbus, Dallas, Denver, Nashville-Davidson, and San Diego, only had a few.

There were also significant regional differences (see Table 3). On average, cities in the South and the West tended to report more measures per service area in their budgets than cities in the East or in the Midwest-Mountain regions did. These regional differences were especially clear in community and neighborhood development and planning and zoning, but less obvious in street services, and fire and emergency medical services.

How Do Cities Report Performance Measures?

Since performance measurement is often linked to strategic planning and goal setting,[22] we expect that cities that had performance measures would also have mission and goal

22. See Cheryle A. Broom and Lynne A. McGuire, "Performance-Based Government Models: Building a Track Record," *Public Budgeting & Finance* 15, no. 4 (1995): 3–17; Cheryle Broom et al., *Performance*

TABLE 3
Average Number of Performance Measures Reported, by Programs and Regions

Program Areas	East	Midwest & Mountain	South	West
Community and neighborhood development	35.0 ($n = 2$)	17.7 ($n = 3$)	78.0 ($n = 8$)	69.0 ($n = 5$)
Economic development	23.5 ($n = 2$)	7.3 ($n = 3$)	15.7 ($n = 7$)	13.2 ($n = 5$)
Fire and emergency medical services	32.0 ($n = 2$)	31.8 ($n = 4$)	48.2 ($n = 9$)	32.7 ($n = 6$)
Library	14.7 ($n = 3$)	19.3 ($n = 3$)	24.9 ($n = 7$)	25.2 ($n = 5$)
Parks and recreation	16.5 ($n = 2$)	35.0 ($n = 3$)	59.8 ($n = 9$)	21.5 ($n = 6$)
Planning and zoning	14.0 ($n = 1$)	12.3 ($n = 4$)	25.8 ($n = 6$)	24.0 ($n = 5$)
Police	15.5 ($n = 2$)	42.3 ($n = 4$)	49.9 ($n = 9$)	39.7 ($n = 6$)
Street maintenance and repairs	18.5 ($n = 2$)	4.0 ($n = 3$)	13.1 ($n = 9$)	10.0 ($n = 6$)

Note: Cities in the East include Boston and Washington, DC. Cities in the Midwest-Mountain region include Columbus, Denver, Detroit, and Milwaukee. Cities in the South include Austin, Dallas, El Paso, Fort Worth, Houston, Memphis, Nashville-Davidson, Oklahoma City, and San Antonio. Cities in the West include Phoenix, Portland, San Diego, San Francisco, San Jose, and Tucson.

statements. Our results confirm this finding (see Table 4). Even some of the cities that did not integrate performance measures in their budgets had a mission statement and a few strategic goals for individual programs. Most of these goals are general statements, such as "responding to citizen requests in a timely and professional manner," and "reducing crime and enhancing the safety of the community."

About 80 percent of the departments or programs also provided specific performance targets, such as "reducing serious collisions by 15%," "maintaining a 11% conviction rate for fires resulting from arson," or "ensuring that 90% of all emergency medical service responses arrive within 8 min from time of dispatch." The cities of Austin, Boston, and Dallas were especially comprehensive and specific in establishing performance targets, and tied program goals, performance targets, and performance results closely in their budget reports.

In addition, 70 percent of the cities showed a comparison of performance results over time. Even among the cities that did not include such a comparison, many often showed an intent to adopt this practice in the near future by presenting a comparison table without actual data, or stating that past data were unavailable but the city had begun to collect the data.

(footnote Continued)
Measurement: Concepts and Techniques, 3rd ed. (Washington, DC: Center for Accountability and Performance, the American Society for Public Administration, 1998); International City/County Management Association, *Comparative Performance Measurement: FY2000 Data Report* (Washington, DC: ICMA, 2001); Robert Lee, "A Quarter Century of State Budgeting Practices," *Public Administration Review* 57, no. 2 (1997): 133–140; Katherine G. Willoughby and Julia E. Melkers, "Implementing PBB: Conflicting Views of Success," *Public Budgeting & Finance* 20, no.1 (2000): 105–120.

TABLE 4
Practices of Performance Reporting, by Service Areas

	Number of cities that report the service	Number of cities with goal statements	%	Number of cities with specific performance targets	%	Number of cities with performance comparison over time	%	Number of cities comparing performance targets and actual results in a particular year	%
Community and neighborhood development	30	28	93	24	80	20	67	15	50
Economic development	28	26	93	22	79	19	68	13	46
Fire and emergency medical services	30	28	93	25	83	20	67	15	50
Library	24	22	92	19	79	17	71	12	50
Parks and recreation	27	26	96	22	81	19	70	15	56
Planning and zoning	30	28	93	24	80	20	67	15	50
Police	29	27	93	23	79	20	69	14	48
Street maintenance and repairs	27	25	93	21	78	20	74	12	44

What is disappointing, however, is that while many cities had specific performance targets, only about 50 percent of the programs showed a comparison between performance targets and actual results in different years. As a result, readers of the performance information cannot judge easily whether the cities had accomplished their program goals or whether there was a performance gap that required remedial action. In addition, almost all cities failed to compare their program performance among different subjurisdictions of a city, or externally with the performance of other similar cities.[23] This was probably because of the fact that many cities did not collect performance data by geographical area or participate in a comparative performance project, such as the one offered by the ICMA, that would allow them to compare their performance with similar jurisdictions.

Types of Performance Measures Reported

Among the 4,817 measures analyzed in this study, 29 percent were outcome or intermediate outcome measures (see Table 5). Although output measures still occupied the biggest share of the total number of measures, the share was only 53.7 percent, which was significantly lower than the level found by an earlier study by the GFOA.[24] Since this study adopted a more "conservative" approach by counting certain measures such as the number of participants or users as "output" measures rather than "intermediate" outcomes, the shift toward outcome measurement could have been shown even more strongly if a different classification method was used.

Outcome measurement was especially common in economic development and police protection, which had about 40–50 percent of all measures in the outcome or intermediate outcome categories. About 37 percent of the performance measures in fire and emergency medical services also belonged to this category (see Table 5). The services that were least likely to use outcome measurement were parks and recreation programs, planning and zoning, and library services.

Most programs did not report many efficiency measures. This finding is consistent with some earlier findings that relied on self-reporting surveys.[25] Among the 4,817

23. There were a few exceptions. For example, San Antonio showed a comparison of the response time of different police departments in the region.

24. In 1994, the GFOA examined the practice of performance measure reporting in operating budget documents. They screened 554 budgets from jurisdictions of varying sizes submitted to the GFOAs Distinguished Budget Presentation Awards Program, and found that of the 554 budgets screened, 330 documents (60 percent) included performance measures. Using a sample of 43 budgets that contained performance measures, they found that the most commonly found measures were output measures, which made up an average of 70 percent of the total number of measures.

25. Patria de Lancer Julnes and Marc Holzer, "Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation," *Public Administration Review* 61, no. 6 (2001): 693–708; Alfred Tat-Kei Ho, "The Impact of Performance Measurement: The Perspective of City Mayors," paper at the Annual Conference of the American Society for Public Administration (Washington, DC: March 15–18, 2003).

TABLE 5
Number of Measures Analyzed, by Types of Measures and Service Areas

	Demand	Efficiency	Explanatory	Input	Intermediate outcome	Outcome	Output	Total
Community and neighborhood development	22	113	1	83	155	115	603	1,092
	2.0%	10.3%	0.1%	7.6%	14.2%	10.5%	55.2%	100.0%
Economic development	1	14	2	13	21	89	105	245
	0.4%	5.7%	0.8%	5.3%	8.6%	36.3%	42.9%	100.0%
Fire and emergency medical services	8	56	3	69	209	91	385	821
	1.0%	6.8%	0.4%	8.4%	25.5%	11.1%	46.9%	100.0%
Library	2	36	—	35	49	12	268	402
	0.5%	9.0%	0.0%	8.7%	12.2%	3.0%	66.7%	100.0%
Parks and recreation	7	51	—	41	80	38	588	805
	0.9%	6.3%	0.0%	5.1%	9.9%	4.7%	73.0%	100.0%
Planning and zoning	8	83	4	28	33	7	175	338
	2.4%	24.6%	1.2%	8.3%	9.8%	2.1%	51.8%	100.0%
Police	3	56	4	58	245	188	333	887
	0.3%	6.3%	0.5%	6.5%	27.6%	21.2%	37.5%	100.0%
Street repairs	3	22	—	8	60	3	131	227
	1.3%	9.7%	0.0%	3.5%	26.4%	1.3%	57.7%	100.0%
All programs	54	431	14	335	852	543	2,588	4,817
	1.1%	8.9%	0.3%	7.0%	17.7%	11.3%	53.7%	100.0%

measures examined, only 10 percent were efficiency measures. A possible explanation for this finding is that many program managers in city governments still lack a clearly defined methodology and the technical capacity to account for costs and analyze program efficiency. In addition, the rising emphasis on results-oriented management and customer satisfaction may have led many city governments to downplay the traditional focus on economy and cost-efficiency in performance measurement and encourage more reporting of outcome and intermediate measures.

Similar to an earlier finding by the 1994 GFOA report, other performance-related measures, such as input, demand, and explanatory information, continued to receive very little coverage in city budgets.[26] Only about 1 percent of the total number of measures were demand or explanatory measures, and only an additional 7 percent were input measures. However, this finding does not imply that cities have totally ignored these types of information for budgetary decision-making. Even though specific input or explanatory measures were not reported, similar information was often embedded in program descriptions or in management discussion of the budget.

Regression Analysis

A regression analysis was conducted to analyze the factors that influenced the types of measures reported in a city budget. The dependent variable is the percentage of different types of performance measures reported in individual programs by the largest 30 cities ($n = 225$). For example, if a city chose not to report any outcome measure, the city has a zero for the percentage of outcome measure. If a city did not report any performance measure at all, it has zeros for all dependent variables. We test whether professional organizations, such as the ICMA and GFOA, had any significant impact on the pattern of performance reporting in budget documents. We expect that the comparative performance program by the ICMA and the budget award program by the GFOA should have a significant and positive impact on the percentage of outcome-oriented and efficiency measures, because both organizations are supportive of performance measurement and have been encouraging their members to adopt performance budgeting techniques. Both organizations are also part of the National Advisory Council on State and Local Budgeting (NACSLB), which embrace outcome measurement and reporting in budget documents.

We also hypothesize that public reporting of performance measurement results on the Internet should have a positive and significant correlation with the ratio of outcome-oriented measures because the public are more interested in these measures than in output or input measures.[27] Moreover, having a professional manager should have a positive impact on the percentage of outcome-oriented measures and efficiency measures,

26. GFOA, 1994.

27. Alfred Tat-Kei Ho and Paul Coates, "Citizen-Initiated Performance Assessment—The Initial Iowa Experience," *Public Performance & Management Review* 27, no. 3 (2004): 29–50.

since professional managers are more likely to embrace the idea of performance measurement advocated by professional organizations and act as a "change agent" to help departments implement the tool.

The regression analysis also includes several control variables. The descriptive statistics in Table 5 show that police, fire-EMS, and economic development programs had a relatively higher percentage of outcome-oriented measures, planning and zoning services had a relatively higher percentage of efficiency measures, and parks and recreational services and library services had a relatively higher percentage of output measures and input measures, respectively. Hence, we use dichotomous dummy variables to control for the unique characteristics of these specific programs.

We also control for the population effect and hypothesize that it has a negative impact on outcome measurement. Larger cities tend to have a less controllable operating environment for public service delivery. As a result, city staff may have less political incentive to report outcome measures publicly, for fear of potential embarrassment if the performance results tend to fluctuate significantly over time. However, we recognize that larger cities usually have greater fiscal and personnel capacity to conduct program evaluation and report outcome measures than smaller cities do. Also, officials in smaller jurisdictions who have more direct contacts with citizens may be more sensitive to public reaction to performance measurement results and may become even less willing to report outcome measures. Therefore, it is possible that the empirical results may contradict our hypothesis and show a positive relationship between population and the percentage of outcome measures.[28]

Table 6 reports the regression results.[29] As expected, the control variables for different service areas in all four regressions are significant, confirming the hypothesis that certain programs, such as police and fire-EMS services, are likely to have certain types of performance measures, such as outcome or intermediate outcome measures, simply because of the nature of the services and long-established professional practices.

The presence of professional managers has a highly significant and positive impact on the adoption and practice of performance measurement, which is consistent with some of the past findings about the roles of managers.[30] They are not only a primary consumer of performance information, but also an important facilitator that encourages organizational learning and adoption of performance measurement by departments and elected

28. We thank one of the anonymous reviewers for this suggestion.

29. In addition to the variables shown in Table 6, we have tested other control variables, such as the size of a program budget, the practice of reporting performance targets, and the practice of comparing performance targets and actual performance. None of them are significant and are therefore not reported.

30. Berman and Wang, 2000; Julnes and Holzer, 2001; Mark D. Rivera, Gregory Streib, and Katherine G. Willoughby, "Reinventing Government in Council-Manager Cities—Examining the Role of City Managers," *Public Performance & Management Review* 24, no. 2 (2000): 121–132; Xiaohu Wang, "Performance Measurement in Budgeting: A Study of County Governments," *Public Budgeting & Finance* 20, no. 3 (2000): 102–118.

TABLE 6

Regression Analysis of the Types of Performance Measures Reported (Dependent variable: Percentage of a specified type of measures, by programs of the largest 30 cities)

	% of outcome or intermediate-outcome measures	% of efficiency measures	% of output measures	% of input measures
Intercept	139.66 (3.70)***	−1.10 (−0.06)	−11.67 (−0.23)	14.60 (1.43)
Participate in the ICMA project (Yes = 1, No = 0)	7.33 (1.67)*	0.26 (0.12)	−7.94 (−1.36)	−0.80 (−0.68)
Number of years with GFOA budget award	−0.47 (−1.81)*	−0.21 (−1.62)	1.66 (4.82)***	0.21 (2.99)***
log(population)	−9.36 (−3.57)***	0.20 (0.15)	0.83 (0.24)	−1.01 (−1.42)
Have a city manager (Yes = 1, No = 0)	14.01 (3.54)***	7.18 (3.57)***	15.53 (2.95)***	0.67 (0.63)
Report performance on the Web (Yes = 1, No = 0)	−3.94 (−1.11)	3.69 (2.04)**	27.68 (5.84)***	0.30 (0.31)
Dummy variable for police services (Yes = 1, No = 0)	30.05 (6.79)***			
Dummy variable for fire-EMS services (Yes = 1, No = 0)	15.32 (3.46)***			
Dummy variable for economic development programs (Yes = 1, No = 0)	17.56 (3.86)***			
Dummy variable for planning-zoning services (Yes = 1, No = 0)		7.40 (3.38)***		
Dummy variable for library services (Yes = 1, No = 0)			17.88 (2.81)***	2.84 (2.2)**
Dummy variable for parks and recreation services (Yes = 1, No = 0)			11.15 (1.85)*	2.05 (1.75)*
R^2	0.34	0.15	0.33	0.10
Adjusted R^2	0.32	0.12	0.31	0.08
Number of observations	225	225	225	225

Note: *t*-statistics in parentheses.
***$P<0.01$; **$P<0.05$; *$P<0.1$.

officials.[31] Hence, having a professional manager increases the likelihood that a city may report performance measures, whether outcome, efficiency, or output measures, in its budget.

The ICMA performance measurement program also has a positive impact on the practice of outcome measurement. Participating in the program, on average, raises the ratio of outcome measures reported in a program's budget by about 7 percent.[32] The GFOA award program, however, has a surprising negative impact, even though the substantive impact is relatively minor. After controlling for other factors, the longer that a city receives the budget award, the less likely it is that the city will report many outcome-oriented and efficiency measures, and the more likely that it will report more output and input measures. While the results seem to be counterintuitive, there are several possible explanations. First, the GFOA does not require reporting of performance measures in its budget award program. Hence, whether a city reports performance measures and what types of performance measures are reported in a budget do not increase the chance of getting the GFOA recognition. In addition, a past study has found that the selection and usage of performance measures tends to be highly stable over time.[33] Because performance measures are not expected by the GFOA, a city that continues to receive the national honor for its budgetary practice year after year has very little incentive to change its reporting format and content. As a result, many GFOA-recognized cities tend to report traditional indicators, such as input, output, and workload measures, rather than outcome and efficiency measures.

The impact of Internet reporting of performance measurement results is also contradictory to our hypothesis—it reduces the percentage of outcome-oriented measures and increases the percentage of output measures. A possible explanation is that since the Internet is a highly cost-effective platform for citizens to access public information, many cities use the World Wide Web (WWW) to show "what they have done for the public." Outcomes, on the other hand, are often difficult to measure and are less amenable to control by departments. Fearing that there may be political embarrassment if program outcomes are below the public expectation or fluctuate significantly over time, cities with Internet reporting have less incentive to report outcomes and instead, tend to focus on output.

Population generally has no significant impact on the choice of performance measures, which is consistent with an earlier finding in county governments.[34] The only exception is that population size is negatively related to the percentage of outcome

31. Ron Kluvers, "An Analysis of Introducing Program Budgeting in Local Government," *Public Budgeting & Finance* 21, no. 2 (2001): 29–45.

32. In a separate regression that only analyzes cities reporting performance measures and excludes those that did not, the positive impact of the ICMA program is even stronger, with an average increase of about 10 percent in the ratio of outcome measures, and the impact is statistically significant at the 5 percent level.

33. Daniel E. O'Toole and Brian Stipak, "Productivity Trends in Local Government Budgeting," *Public Performance & Management Review* 26, no. 2 (2002): 190–203.

34. Wang, 2002.

measurement, and that this impact is significant. On average, a 1 percent increase in population leads to a 9 percent reduction in the percentage of outcome measures. The finding seems to support our hypothesis that given the differences in the political environment, larger cities are generally more risk-averse and are less likely to adopt outcome performance reporting than many medium-sized cities are.

DISCUSSION

The results above show that significant progress has been made in performance measurement and reporting over the past few decades. City officials have not only expanded the scope of the practice by reporting measures for almost all city services, including areas that have many technical measures, such as law enforcement, public works and city planning, and areas that are less technical and are more customer-oriented, such as library and parks and recreation programs, but have also expanded the depth and sophistication of performance measurement. Besides traditional performance measures that are workload- and output-focused, many cities also report outcome and intermediate outcome measures. Even in highly technical areas such as street repair and maintenance, many city officials embrace outcome reporting by using intermediate outcomes measures, such as results of citizen satisfaction surveys and complaint statistics.

Only a few municipal services, including planning and zoning, parks and recreation, and libraries, remain primarily output-focused. The pattern is probably caused by the inherent nature of these services. Without carefully designed program evaluations and studies over a long period of time, officials in these areas often have great difficulty isolating the influence of environmental factors and demonstrating the link between their service efforts and the social and economic impact of their programs. As a result, many city officials in these areas prefer to focus on output measures and document what activities they have done.

This does not mean that there is no room for greater outcome focus in these services. For example, managers in these programs may report more measures related to responsiveness to citizens' requests or complaints, especially because many of their activities are of a customer-service-oriented nature. Moreover, these managers may report more intermediate outcome measures, such as users' evaluation of service quality and results. Finally, they may conduct occasional evaluations of some of the service outcomes. The city of San Diego, for example, reported the percentage of adult learners showing skill improvement by a survey. The city of San Jose also reported the percentage of literacy program users who showed improvement in reading skills and the percentage of parents and caregivers who reported that they had read more to their children following a preschool reading readiness program. These precedents of reporting outcomes show that stronger outcome orientation is possible in these areas, but it may require more internal capacity building for program evaluation and more careful thinking by managers in the execution process.

Our results above also show an encouraging sign that many cities have made attempts to link performance measurement with goal setting, strategic planning, and budgeting by reporting performance measures along with mission and goal statements in budget documents. However, we also notice that in some cities' budgets, some of the performance goals were too broad, some of the performance targets were not quantifiable or specified clearly, and the link between program goals and performance targets was often too vague. Hence, greater attention to these details is needed in the future to improve the quality of performance reporting. Merely reporting many goals with irrelevant performance data does not contribute to "results-oriented" thinking and cannot help decision-makers see whether a program has really accomplished what it is supposed to. Information overload may in fact hurt the credibility of performance measurement and discourage decision-makers from paying attention to the measures.

The finding that most cities do not disaggregate performance information, especially outcome or intermediate outcome measures, by geographical area is also a concern that should be addressed in the future. One earlier study has suggested that citizens and elected officials are often interested in knowing how public services perform in different local areas, such as in neighborhoods or voting precincts.[35] For example, they want to compare the response time of the police and fire in different subdivisions of the city, or want to know the results of citizen satisfaction surveys by neighborhoods to see if there is any inequity in service delivery. To make performance measurement meaningful to the public and generate the necessary political interest and demand for the practice, citywide statistics of performance results are not sufficient. City officials need to take an extra step in analyzing performance data and report the results at multiple geographical levels so that the information is more relevant and politically valuable in the budgetary and policymaking process.

CONCLUSION

For the past decade, advocates of performance measurement and budgeting have emphasized "result" or "outcome" orientation in using and reporting performance measures. However, very few systematic studies have been done to analyze whether this approach has been adopted by public agencies, and even fewer studies have examined this question objectively and systematically by nonsurvey methods. Using the budgetary documents of the largest cities of the U.S., this paper analyzes more than 4,800 measures and shows that significant progress toward outcome measurement has been made. Most of these cities have elaborated performance measurement to show how they use taxpayers' money to accomplish results. In addition, many of them have integrated performance measurement and reporting with strategic planning and goal-setting. Even though there is still room for future improvement, the results are encouraging and

35. Ho and Coates, 2004.

suggest that many cities have made considerable attempts to enhance public accountability and results-oriented management.

While this study has offered new insights about the practice of performance reporting in U.S. municipalities, there are many questions that remain unanswered. For example, it is unclear how the improvement in performance reporting has impacted the quality of decision-making. Do city council members feel more informed? Have managers become more outcome-oriented and customer-focused? How do cities integrate performance reporting with other budgetary and decision-making mechanisms to ensure that performance measures are used effectively to improve service delivery? These are some of the questions that should be examined more carefully by future research.