

# **PERFORMANCE MEASUREMENT, ACCOUNTABILITY, AND TRANSPARENCY OF BUDGETS AND FINANCIAL REPORTS**

**JAMES MELITSKI**

Marist College

**AROON MANOHARAN**

Kent State University

## **ABSTRACT**

In an age of global public cutbacks and talks of austerity, public managers are wary of publishing performance data online. Yet, public organizations are using information technology to measure and manage their performance now more than ever. At the same time, communities that engage citizens online and report performance information are more accountable to the public and build public trust. This paper examines the link between performance reporting and financial decision-making by analyzing award winning financial reports and budget reports identified by the Government Finance Officers Association in 2009. Drawing on the Governmental Accounting Standards Board's (GASB) criteria for performance reporting, a rubric is applied to both public budget and public financial reports. Second, a data dictionary or lexicon for performance measurement is developed and applied to the same documents to determine the extent to which the documents address performance measurement. Lastly, the relevance of performance reporting is examined in both planned and actual expenditures by comparing the budget and financial reports.

**Keywords:** Performance reporting, budget reports, performance measurement, e-government, public reporting.

## **INTRODUCTION**

The link between organizational performance and policy decisions that allocate resources are often difficult to assess. Public administrators understand that evaluating the performance of public agencies is difficult to measure quantifiably, and the incremental nature of public policy, as well as the budgeting

processes makes the link between the evaluation of agency performance and the allocation of future resources difficult. Further, public administration scholars have long struggled with chicken-and-egg dilemmas of social conditions and government expenditures. Do crime-rates provide a rationale for police spending? Does the employment rate reflect the performance of departments of labor? In other words, is it appropriate or even possible to apply rational public policy models to the outcome measurement process that reflects both agency performance and broader socio-economic conditions?

The discussion is further complicated in times of economic and social distress, as outcome indicators such as unemployment and income levels consistently decline. A recent review of performance data published by the City of New York indicates that 204 of the city's 519 performance indicators are currently declining (<http://www.nyc.gov/html/ops/cpr/html/home/home.shtml>). In an era when socio-economic outcome indicators are in decline, how are policy makers to integrate the information into their decision making process. What is the appropriate level of resource allocation when outcome indicators spanning multiple agencies decline?

Such dilemmas inhibit public officials from effectively integrating objective performance data into public budget and financial decisions on a regular basis. However, an inability to bring performance information into the financial planning and analyses processes should not stop public administrators from attempting it. Without data on the performance of their agencies, policy makers cannot effectively make informed decisions, rendering any hope of continuous improvement impractical. Furthermore, accurate, timely and historically comparable performance data is vital to managers as they seek to hold their agencies accountable to the public.

This paper assesses budget and financial reports in U.S. states to determine the extent to which they integrate performance data into their publicly available reports. First, we examine relevant literature in performance reporting. The paper then assesses budgets and popular annual financial reports in U.S. states to determine the extent to which they integrate

performance data. We employ two complementary methodologies to conduct the content analysis. Drawing on the Governmental Accounting Standards Board's (GASB) criteria for performance reporting, a rubric is applied to both public budget and public financial reports. Second, a data dictionary or lexicon for performance measurement is developed and applied to the same documents to determine the extent to which the documents address performance measurement. Lastly, the relevance of performance measurement is examined in both planned and actual expenditures by assessing budget and financial reports.

### **LITERATURE REVIEW**

Transparency, responsiveness and citizen engagement are common themes among public management scholars in performance measurement, e-government, as well as, public budget and finance. Open government has become a common trope of recent political actions; the public, especially in the remnants of the Great Recession, demands greater understanding of where their tax dollars go. Polls conducted by Gallup from 2001 to 2012 reveal that over 65% of the population considers economic issues to be the most important problem facing the United States (<http://www.gallup.com/poll/1675/most-important-problem.aspx>). Until recently, few recognized the connection between technology, finance and performance, which all espouse the similar goals of engaging citizens and allowing policy makers to hold agencies accountable for productivity improvement (Justice et al., 2006). Indeed, there is a paradox associated with using outcome data as a basis for making decisions about future resource allocations. As Kasdin (2010) points out, rational responses to declines in outcome measures can involve both increases and decreases in agency funding levels. For example, if job growth declines, a rational response for policy makers is to increase funding for the Department of Labor due to need. Yet, a second rational response for decreasing funding because the agency is not fulfilling its mission is also reasonable. In sum, reasoned arguments can be made for both increasing and decreasing agencies' funding based

on poor performance. As an alternative, some scholars argue that the more appropriate use for performance information is to inform the budgeting process rather than mandate particular decisions (Aristigueta & Justice, 2006). In other words, performance measurement is best used as a management tool as opposed to a method for allocating budgets (Joyce, 1993; Melkers & Willoughby, 1998).

Performance measurement advocates suggest that building systems for reporting objective metrics is an essential component of public management (Ammons, 1995; de Lancer Julnes & Holzer, 2001; Poister, 2010; Poister & Streib, 1999; Wholey & Hatry, 1992; Yang & Holzer, 2006). Movements to link budgeting and performance have arisen since the Hoover commission in 1947 (Howard, 1973; Rubin, 1997). In the 1960s and 1970s federal efforts were difficult to sustain. Budget reforms like zero based budgeting gained popular favor and performance budgeting was viewed as difficult to quantify and were unwieldy time consuming (Jordan & Hackbart, 1999; Lee & Johnson, 1994).

In the 1990s, performance budgeting was rediscovered as a part of the reinventing government movement and the Government Performance and Results Act (GPRA) of 1993. GPRA requires federal agencies to submit performance data along with budget requests and reinvigorated the performance budgeting movement (Jordan & Hackbart, 1999; Roth, 1992). As a result, global reinvention efforts in the field of public administration have included both performance measurement and performance budgeting (Gilmour & Lewis, 2006; Kettl, 2000).

At the state level, performance budgeting has become a common tool for policy makers. For example, a study in 1980s found performance measurement a useful and popular tool in 31 of the 50 U.S. states (Botner, 1985). However, several state level analyses failed to demonstrate a link between performance measurement, spending, staffing, and valid metrics (Connelly & Tompkins, 1989; Jordan & Hackbart, 1999; Lauth, 1985). Despite concerns about direct applicability of performance measurement in the budgeting process, performance budgeting

has been widely adopted (Gilmour & Lewis, 2006; Schick, 1990).

In 1998, Melkers and Willoughby began studying states with legislated performance measurement mandates and found that 47 of 50 states had some form of performance budgeting in place (Melkers & Willoughby, 1999, 2001). By 2009, a study of U.S. States indicates that 39 states integrated performance systems into their public budgeting processes (Lu, Willoughby & Arnett, 2009). Of the 39 states with integrated performance systems, 19 have performance portals or centralized online depositories for performance data (Yi & Willoughby, 2011). It is also worthy of note that public organizations are using the Internet and social media technologies as a means for engaging citizens and improving trust by making government more transparent and reporting on their performance, budgets and expenditures.

A significant factor in determining the success of performance budgeting is the performance measures/information itself. In a study on performance information in 1984, MacManus reported that more than 80 percent of public performance reports did not provide measures of effectiveness or efficiency. According to her, without proper linking between inputs and outputs, any changes in budgets would lack justification and are influenced by political favoritism. Based on a survey of 205 city budget directors, Moore found that more than 76 percent of the budget directors considered lack of performance information as a problem with more than half of them considering it as a serious problem (Grizzle, 1986). More recently, Ho & Ni (2005) found many of the largest cities in the United States reporting both outcome and output measures online, as well as integrating performance measurement and reporting with strategic planning and goal-setting.

In addition to the performance measurement and budgeting efforts, the performance reporting movement in the United States has its roots in the public reporting efforts, which began during the early half of the twentieth century. The New York Bureau of Municipal Research is recognized as an early exemplar in making public information accessible to the public in understandable formats (Jones, Scott, Kimbro & Ingram,

1985; Kahn, 1997). Public reporting is an act by the government to ensure an informed citizenry. According to Lee, public reporting relates to managerial initiatives for systematic transparency covering and regularly informing the public about government operations, thereby promoting an informed citizenry (2004a). Gradually, reporting began to assume a narrower role of informing the public on the performance of the public sector agencies and was considered necessary for the government to be accountable to the public. Lee states that early public reporting movements justified reporting on the basis that transparency, accountability, and effective management are facilitated by public reporting and these movements were seized upon as effective strategies and an essential foundation for good government in democratic societies (2004a).

Yet, as the public reporting movement matured, it went beyond an accountability tool to the public; as some scholars during this period also hinted at public reporting as a tool of control. Further, performance measurement falls into Dubnick's (2005) accountability paradox as it represents both a tool for quantifiably measuring manager performance toward the achievement of program outputs, while also addressing the larger qualitative issue of ensuring that public agencies serve the public interest. As a result, reporting of public performance seeks to achieve the goal of enhancing public trust in a time of sustained decline. Additionally, some ambitious candidates for public office seize upon public reports as a means of critiquing incumbents, threatening executive sponsorship of public reporting. As a result, public agencies are often reluctant to publish reports that showed any declines in the performance of public agencies. Recently, however, the growing emphasis of citizen participation and citizen engagement has rejuvenated the phenomenon of public performance reporting in the United States (Caddy & Vergez, 2004).

Isenmann, Bey & Welter (2007) argue that online reporting allows stakeholders to engage in a rich, personalized, and sophisticated interaction with information that facilitates greater understanding of the data. Online reporting also improves the quality of information and makes it more accessible to citizens (Holliday & Kwok, 2004; Scavo & Shi,

1999). Applying e-government literature to performance reporting indicates that online reporting can inform managerial decisions, help frame policy deliberations and engage the public (Chadwick & May, 2003). Moreover, Holzer et al. (2004) hypothesize that online interactions between citizens and government improves transparency and builds public trust.

In addition, online reporting leverages government technology and enables effective communication and reporting to citizens by allowing ubiquitous access 24 hours a day, 7 days a week. Often, governments utilize their websites to publish results of their performance measurement systems in the hopes of strengthening accountability through transparency. Lee defines e-reporting as “the administrative activity that uses electronic government technology for digital delivery of public reports that are largely based on performance information. E-reporting is a tool of e-democracy that conveys systematically and regularly information about government operations that is valuable to the public at large, in order to promote an informed citizenry in a democracy and accountability to public opinion” (2004b, p. 11).

Public reporting of performance measurement data has its own distinct history. In 1994, the Governmental Accounting Standards Board (GASB) released the “Concepts Statement No.2 on Service Efforts and Accomplishments (SEA) Reporting,” stating that performance information needs inclusion in general purpose external financial statements. The recommendation further emphasizes the need to link performance measures with financial reporting. A significant statement by GASB, the report established the need for linking performance information to the allocation of resources, and it institutionalized performance measurement and reporting as an integral part of the financial reporting process (GASB, 2003). To achieve this objective, GASB proposed three broad categories and outlined sixteen criteria to guide the development of external performance reports, based on a series of discussions among managers, academics and other practitioners. The first such report, *Reporting Performance Information: Suggested Criteria for Effective Communication (2003)* presented various techniques to convey the efforts, challenges, approaches, and successes

associated with delivering public services.

GASB standards established three goals for performance reports: First, GASB requires an “External Report on Performance Information.” In essence, this requirement stipulates that performance reports need clear organization. To measure this goal, the report lays out criteria 1-7, which examine the purpose and scope of the report, and whether major goals are stated and established by key stakeholders. Meeting this goal requires multiple levels of reporting, and entails a macro-level analysis of results, challenges and key strategic performance measures.

The second GASB goal for performance reporting outlines “what performance information to report” in criteria 8-14. To fulfill this goal, reports must conduct a detailed examination of performance information given in a report. For example, these criteria require that performance reports measure relevant metrics, examine resources used, efficiency, citizen perspectives, the availability of comparison data, factors that might affect the results, whether the information is aggregated or disaggregated and the consistency of the methodology used to collect the performance data.

Finally, GASB examines how well the performance data is communicated in its final category dubbed, “communication of performance information” which is measured in the final two criteria (15 and 16). These criteria measure accessibility and clarity of the report as well as whether the information in the report is collected regularly. Figure 1, below, lists a summary of the 16 GASB criteria for preparing performance reports.



Figure 1  
*GASB Criteria for Performance Reporting*

<p><b>1. Purpose and Scope:</b> Highlights what the report aims to communicate to users along with a listing of the programs and departments that are covered in the report.</p>	<p><b>9. Resources Used and Efficiency:</b> Determines if the financial input into the program is linked to the output of the services in terms of effectiveness and efficiency measures.</p>
<p><b>2. Statement of Major Goals and Objectives:</b> Identifies the goals and the objectives of the report and explains the source of these goals and objectives.</p>	<p><b>10. Citizen and Customer Perspective:</b> Examines citizens' opinion/perception with the performance results and compares it to that of the management of the agency.</p>
<p><b>3. Establishing Goals and Objectives:</b> Determines if the stakeholders who are involved in selecting the goals and objectives are listed and the extent of their involvement.</p>	<p><b>11. Comparisons for Assessing Performance:</b> Checks for any comparative analysis with respect to established targets and other departments.</p>
<p><b>4. Multiple Levels of Reporting:</b> Determines if the report is able to guide specific users to their required performance information.</p>	<p><b>12. Factors Affecting Results:</b> Identifies external and internal factors that affect the performance measurement process.</p>
<p><b>5. Analysis of Results and Challenges:</b> Checks if the report contains the management's perspective on the performance results and also discusses the major challenges faced.</p>	<p><b>13. Aggregation and Disaggregation of Information:</b> Ensures that the performance information is aggregated or disaggregated appropriately.</p>
<p><b>6. Focus on Key Measures:</b> Ensures that the report identifies key measures of performance and determines how these measures guide citizens to decision making.</p>	<p><b>14. Consistency:</b> Studies the methodology utilized in collecting the performance results and checks for changes introduced in the measures.</p>
<p><b>7. Reliable Information:</b> Checks the reliability and relevancy of the performance data.</p>	<p><b>15. Easy to Find, Access and Understand:</b> Measures the availability, accessibility of performance reports and how it can be identified.</p>
<p><b>8. Relevant Measures of Results:</b> Determines the extent to which performance results are linked to the goals, statements of the program in the report.</p>	<p><b>16. Regular and Timely Reporting:</b> Identifies the time period of performance reporting.</p>

For the purposes of this research, we re-name the three GASB goals as follows: 1. Organization of Report (1-7); 2. Performance Information (8-14); and 3. Communication and Contact Information (15-16). We assess the three goal categories

and evaluate the 16 criteria using a rubric for assessing the extent to which different reports address performance information. By using these criteria, governments in the United States and abroad can utilize information and communication tools to the full potential in providing public information and services online to their residents. Besides the promise of improving the efficiency and effectiveness of public service delivery, performance reports available online draw on the benefits of e-government, which espouses to enhance government accountability by expanding public access to information.

### **RESEARCH METHODOLOGY AND ANALYSIS**

To assess the impact of public performance data on budget and financial decisions, this research examines popular annual financial reports (PAFRs) and budget reports from U.S. states and assesses the extent to which they address performance. A purposive sample was identified using the Government Financial Officers Association's annual award program for exemplar budgets and PAFRs. As a nonprobability sample, the documents represent best practices, and as such, it is expected that they exemplify performance reporting in budgeting and popular annual financial reports. The purposive sampling technique, analyzed 17 documents (budget reports and PAFRs) identified as exemplars by the Government Financial Officers Association (GFOA). The PAFR Awards Program was established by the GFOA in 1991 to motivate state and local governments to utilize performance information from their comprehensive annual financial reports (CFARs) and publish PAFRs that are easily understandable by the public. The state and local government that are considered for best practices in PAFRs should have either have received the GFOA's Certificate of Achievement for Excellence in Financial Reporting for the comprehensive annual financial report (CAFR) of the previous year or 2) ultimately receive the GFOA's Certificate of Achievement for Excellence in Financial Reporting for the current year. The evaluation process involves five categories with varying weights of importance – reader appeal (10%), understandability (25%), distribution methods (7.5%) and other

(such as creativity, notable achievement) (7.5%) and the remaining 50% for the overall quality and usefulness of the report. Similarly, the Distinguished Budget Presentation Award Program was established by GFOA in 1984 to guide state and local governments in preparing high quality budgets and also involves the guidelines of the National Advisory Council on State and Local Budgeting (GFOA).

In 2010, GFOA identified eight state budgets as exemplars through their “Distinguished Budget Award” program. The budgets were prepared in 2009, and they typically cover the 2010-2011 fiscal year. Budgets analyzed are from the following states: Pennsylvania, Illinois, Tennessee, Massachusetts, Ohio, New Jersey, West Virginia, plus Washington, DC. The second set of documents examined were again identified by GFOA in 2010. GFOA recognized a total of nine state governments for their “Distinguished Popular Annual Financial Report.” The PAFRs were typically prepared in 2009, covering the previous fiscal year (2008-2009). The nine PAFRs analyzed are: Illinois, South Carolina, Nevada, North Carolina, New Hampshire, New York, Idaho, Virginia, plus Washington DC.

Methodologically, two techniques were applied to each of the 17 documents. First, a content analysis was conducted using a rubric that operationalizes the GASB 16-point performance reporting standards. The content analysis employed a double blind review process to ensure reliability. As is customary in qualitative research, when reviewers did not agree on a report’s score, the report was reexamined and a consensus was reached (Miles and Huberman 1984). The performance reporting rubric contains 40 dichotomous measures that were applied to each budget. The resulting analysis evaluates each budget on a scale of 0-40, using a dichotomous scale of 1 and 0, representing the presence or absence of each feature. Next, a lexical analysis was conducted on each document using a dictionary of performance keywords.

The first step in conducting the lexical analysis is developing a data dictionary of performance reporting terms from practitioner articles, public reports and the academic literature. To develop the performance dictionary twenty-two

documents were identified from between 1978 and 2010 (appendix A). The documents were analyzed using the qualitative research software, YoshiKoder, to determine the most commonly used words. The software identified more than 25,000 words used across the 22 performance documents. The authors used an iterative process to identify a list of the 47 most frequently used terms that indicate a document's emphasis on performance.<sup>1</sup>

## RESULTS AND ANALYSIS

After creating the dictionary of performance terms, the relative frequency of performance term usage across the 17 budgets and PAFRs was calculated. The analysis identifies between 15,000 and 20,000 distinct terms in both the budgets and PAFRs. This section assesses the results of the computer generated lexical analysis and the traditional content analysis performed by researchers using a rubric. Below, figure 2 shows the relative frequency of keywords from the performance dictionary among the nine budget documents.

States with the highest usage of performance terms from the performance dictionary in their budget reports were Tennessee and Ohio, which used performance terms of more than 8.7 and 8.4 per thousand words respectively. The lowest was New Jersey, which used a performance term 2.1 times per thousand words. The state of Ohio's *Office of Accountability and Results* coordinated the performance measurement process among the various agencies and enabled tracking of the performance measures through the <http://results.ohio.gov> website. In Ohio, various state agencies also have performance contracts with the Governor, referred to as Flexible Performance Agreements, that require directors to track and report key performance measures. Tennessee uses a slightly different model whereby the individual agencies are expected to submit a strategic plan along with a performance-based budget, based on the guidelines of the Governmental Accountability Act, which are reviewed by the Commissioner of Finance and Accountability.

Figure 2  
*Lexical Analysis of Budget Documents*

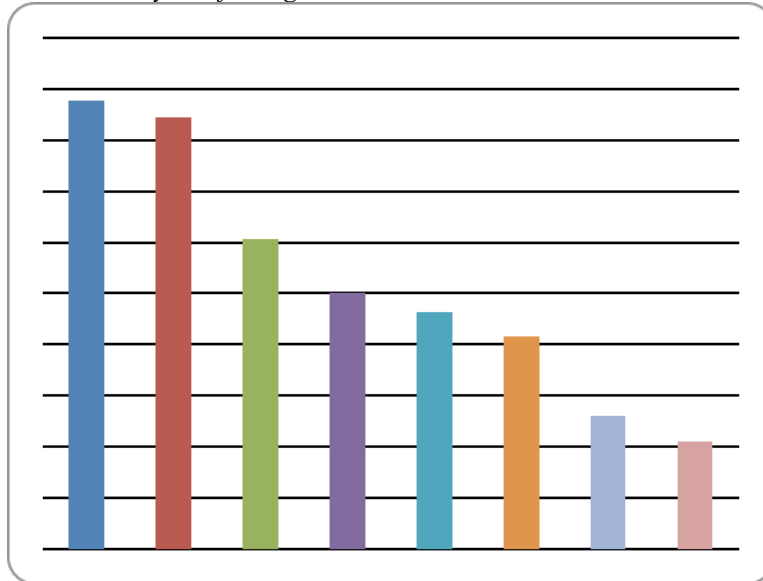


Table 1 below represents the score of each of the ten budget documents analyzed using the performance rubric based on the GASB 16-point criteria. The content analysis results are similar to that of the lexical analysis. The top four states in the lexical analysis (TN, OH, IL, MA) fared well in the traditional content analysis. However, the content analysis shows the budget reports for Pennsylvania and Washington, DC as most comprehensive in their integration of performance measurement into the budget process. Interestingly, Pennsylvania and Washington DC represent the two longest budget reports at 1,072 and 3,907 pages, respectively. By comparison, if we remove these two outliers, the average length of the remaining budget documents is 486 pages per report. Despite the use of relative frequency measures, the size of the Pennsylvania and Washington, DC budgets may prevent the lexical analysis from adequately analyzing their content. This suggests that size may be a factor in conducting the lexical analysis. Excluding Washington DC and Pennsylvania, both the lexical analysis and the traditional content analysis identify West Virginia and New

Jersey as the states that least incorporate performance information. Not surprisingly, these documents were the two smallest budget reports analyzed.

Table 1  
*Content Analysis of Budget Reports*

<b>State</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>Total</b>
DC	Score	13.000	14.000	4.000	31.000
	z-score	1.684	1.397	0.540	1.510
PA	Score	10.000	11.000	4.000	25.000
	z-score	0.561	0.776	0.540	0.722
IL	Score	10.000	10.000	4.000	24.000
	z-score	0.561	0.569	0.540	0.591
TN	Score	8.000	11.000	4.000	23.000
	z-score	-0.187	0.776	0.540	0.460
MA	Score	9.000	5.000	4.000	18.000
	z-score	0.187	-0.466	0.540	-0.197
OH	Score	7.000	4.000	4.000	15.000
	z-score	-0.561	-0.672	0.540	-0.591
NJ	Score	4.000	1.000	3.000	8.000
	z-score	-1.684	-1.293	-1.620	-1.510
WV	Score	7.000	2.000	3.000	12.000
	z-score	-0.561	-1.086	-1.620	-0.985
Average		8.500	7.250	3.750	19.500
Stdev		2.673	4.833	0.463	7.616
Number	of				
Metrics		16	19	5	40

1 – Organization of Report

2 – Performance Information

3 – Communication and Contact

Further, Table 1 shows the greatest variation in the second goal category “Performance Information.” This manifests as two distinct models for incorporating performance information into state budgets. The first descriptive model allows states to describe their performance measurement initiatives, without integrating performance data into the budget.

The second integrative model, involves a traditional performance budget, whereby state budget tables either incorporate specific outcome measures or a separate table of outcome measures is included in the budget following agency budget tables.

Among the exemplars studied in this research, several of the states using the first descriptive model included excellent citizen's guides that discuss the extent to which performance measurement is used in the budget process. However, despite narrative descriptions of the performance measurement process, many states do not amalgamate performance data into their budget reports. As such the first model demonstrates that while performance information may be collected and analyzed, it is not directly integrated into the budgetary decision making process. This led to the high amount of variation in assessing performance information available in the report.

Pursuant to the first model of describing statewide performance measurement processes, several states provide links in their budgets directly to performance portals designed specifically for measuring state agency performance. For example, the Ohio and Massachusetts budget reports both contain links to external performance measurement sites. The Massachusetts performance measurement site, MassGoals, is an excellent performance measurement site; however, there is no way of knowing the extent to which it is used by policy makers to inform budget decisions. The Ohio budget report represents a hybrid approach in that while it includes a link to external performance initiative, the report contains a departmental narrative, which often includes broad highlights and some performance measures.

The second integrated model is a more traditional performance budget, where the budget report contains a separate section for each department including a narrative description of its mission and key goals followed by its budget request for the year. In Illinois, the narrative sections included key strategic objectives, and a few performance measures, but this information was often inconsistent. The reported performance measures compare across years, along with projected performance levels for FY 2010. In a typical performance budget format, specific performance measures, metrics, or benchmarks are reported after

the budget tables. In the best cases such as Washington, DC and Pennsylvania, these tables of performance data include historical data and goals.

It is also worth noting that using the performance-reporting rubric on budgets is problematic. For example, the section that assesses a report's conclusion is difficult to apply to state budgets, which often do not have a formal conclusion. The typical format for a state budget includes a narrative, including a message from the governor, followed by statements of revenue forecasts followed by departmental and capital budget needs for the upcoming year in tabular format. Finally, given the time period of the documents, it is understandable that many of the budgets describe a poor economic climate as an external rationale for potential performance declines in the future. Determining the extent to which these describe outcome measures for state governments is challenging. For example, some states report demographic changes in their regions and related issues such as housing, credit, employment and their potential impact on the state finances. The reports of Pennsylvania and Tennessee discuss the major performance related challenges to the organization in achieving its mission/goals/objective.

While some states use socio-economic shifts as an indicator of public performance, few states discussed internal factors that may impact agency performance in the future. Given the reluctance of policy makers to report poor performance, this is hardly surprising. Finally, consistency of metrics over time is essential to comprehensive performance measurement, yet only the budgets of Pennsylvania and Washington, DC reported any changes in performance measures or methodology.

As our focus on performance measurement shifts from planning of future expenditures to actual expenses incurred, Figure 3 (below) shows the relative frequency of terms from the performance dictionary in 2008 Popular Annual Financial Reports (PAFRs). The keyword analysis shows Illinois as an outlier using performance terms nearly two times (1.97) for every 100 words. Rounding out the top four were Nevada at 5.07 performance terms per thousand, South Carolina at 4.8 performance words per thousand, and New Hampshire, which



used 4.48 key performance words for every thousand total words. Of the best practices identified by GFOA, New York and Virginia used the key performance terms least frequently with 1.88 and 1.06 key performance terms per thousand.

Figure 3  
*Lexical Analysis of PAFRs*

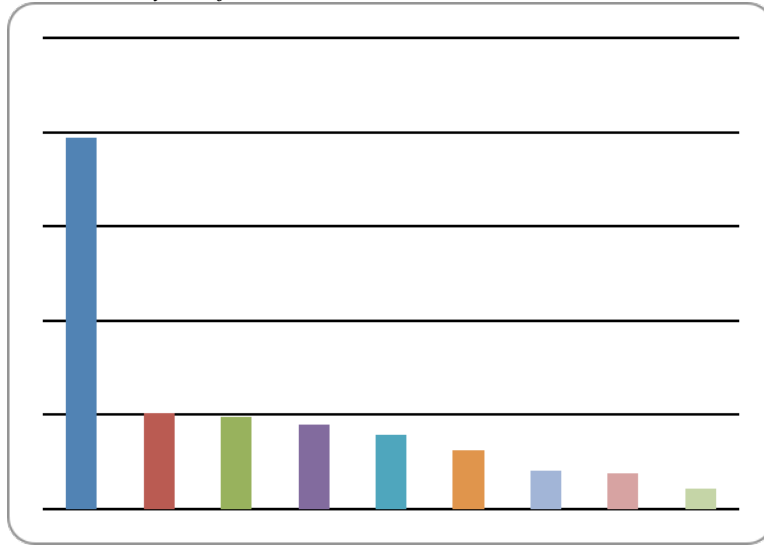


Table 2 (below) reveals that the top four states identified in the traditional content analysis were Illinois (26), South Carolina (24), Nevada (19) and Washington, DC (18). Both the lexical keyword analysis and the traditional content analysis identify Illinois, South Carolina, and Nevada. Again, Washington, DC does not score well in the lexical analysis, while scoring highly in the traditional content analysis. Since the PAFRs represent citizen friendly financial narratives, most are considerably smaller than the budget documents. Again, document size may impact the results, as Illinois represented the longest PAFR at 294 pages. Removing Illinois as an outlier, the average size of the remaining PAFRs was 19.5 pages.

Table 2  
*Content Analysis of PAFRs*

<b>State</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>Total</b>
IL	Score	10.000	12.000	4.000	26.000
	z-score	1.249	1.951	na	1.781
SC	Score	11.000	9.000	4.000	24.000
	z-score	1.699	0.806	na	1.330
NV	Score	6.000	9.000	4.000	19.000
	z-score	-0.550	0.806	na	0.201
DC	Score	7.000	7.000	4.000	18.000
	z-score	-0.100	0.042	na	-0.025
NC	Score	7.000	6.000	4.000	17.000
	z-score	-0.100	-0.339	na	-0.251
NH	Score	8.000	5.000	4.000	17.000
	z-score	0.350	-0.721	na	-0.251
NY	Score	7.000	5.000	4.000	16.000
	z-score	-0.100	-0.721	na	-0.477
ID	Score	4.000	5.000	4.000	13.000
	z-score	-1.449	-0.721	na	-1.154
VA	Score	5.000	4.000	4.000	13.000
	z-score	-0.999	-1.103	na	-1.154
Average		7.222	6.889	4.000	18.111
Stdev		2.224	2.619	0.000	4.428
Number of Metrics		16.000	19.000	5.000	40.000

1 – Organization of Report

2 – Performance Information

3 – Communication and Contact

Despite the outliers (Illinois and Washington, DC), both assessment techniques identified Nevada and South Carolina as states with exemplar PAFRs. Furthermore, both techniques identified Idaho, New York, and Virginia as documents in our sample that address performance measurement the least.

These results suggest several intriguing findings. First, the Illinois PAFR is unique as it was the only document to use

the GASB performance reporting format to structure its PAFR. While the Illinois PAFR is noteworthy for its comprehensiveness, its length suggests that it may not be as citizen friendly as other PAFRs, which are nearly one-tenth its length. The document does provide a good explanation of the GASB's Service Efforts and Accountability (SEA) reporting standards, as part of the Public Accountability Project; however, the PAFR intended to discuss the program information from agencies 'in detail.' Also, many of the PAFRs, such as Virginia, include macro economic data such as employment data. While these are outcome measures for many state agencies, particularly those dealing with labor policies and economic development, they also represent general economic indicators, which assess the external economic environment. The use of such data is necessary but it would be helpful to demonstrate how the outcome measures relate to the performance of various agencies.

In other words, just because a PAFR mentions jobs created in the last fiscal year, it does not mean the PAFR effectively measures the performance of economic development initiatives. As with budget reports, the poor economic climate in 2008 and 2009, when most of the documents were written, allowed many states to discuss broad outcome measures in framing their economic climate in a state. Given the challenges associated with outcome measures identified in the literature review, it is not surprising that states were reluctant to link these external outcome measures with the performance of specific agencies.

In addition, the overall comprehensiveness of the performance information provided in the PAFRs is problematic. The focus of the PAFRs is appropriate financial information, and as a result, the performance information included tends to be cherry-picked to either show improvement or highlight agency accomplishments. In many cases, PAFRs list one table of performance information that, while interesting, is neither strategic nor comprehensive. For example, New Hampshire includes a table that compares the number of bridges and roads in need of repair compared with the number repaired in the previous fiscal year. While this is compelling performance information about the Department of Transportation, it falls short

of a strategic measure of performance for all of state government. Similarly, the Nevada PAFR is worthy of note for including a table of aggregate performance measures that includes historical data and goals for reducing highway fatalities, the percentage of state roads in fair or better condition, and high school graduation rates. While the performance information is informative, there is no indication that the measures are comprehensive or strategic.

Perhaps the most telling assessment in this research involves the comparison between the budgets and PAFRs. Intriguingly, both the lexical and traditional content analysis show a high degree of variation among the reports as indicated in Tables 1 and 2. The content analysis indicates that there was more variation among the PAFRs as compared to the budgets, while the traditional content analysis shows more variation among the budgets than in the PAFRs. Again, this may be attributable to the nature of the two methodologies. Further, the PAFRs tend to be designed with citizens in mind and GFOA has basic criteria for their evaluation. In contrast, there is no uniform format for budgets. As a result, the traditional content analysis detects less variation among the PAFRs. The lexical analysis may be more precise in its ability to detect key words, but does not detect the context provided by the narrative PAFRs, and as such, it identifies the language used in the budgets as more consistent.

As a result, both methodologies identified Illinois as having the PAFR with the most emphasis on performance measurement. Both this structure and the Illinois emphasis on performance in its PAFR, which by nature is designed to make financial data relevant to citizens, is worthy of note. Also intriguing was that the two methodologies yielded very different results for Washington DC. This finding speaks to the reliability of both techniques and more research is needed to examine the validity of the techniques. In particular, both content analysis techniques hold the potential for bias. Intercoder reliability and selection of the keywords must be reexamined in both techniques to ensure consistency and improved reliability.

Table 3 shows that on average the budgets had higher scores than the PAFRs in both the lexical and traditional content

analyses. The lexical analysis shows that the use of performance terms in budgets was 2% greater than in PAFRs; while the traditional content analysis shows that average scores of budgets were 7% higher than PAFRs. One reason for the difference may be that the standards for including performance information in PAFRs are less well-known. For e.g., examples of performance budgets are available, while there are fewer examples for adapting comprehensive annual financial reports into citizen friendly PAFRs that include strategic performance measures.

Table 3

*Average Scores for Lexical and Traditional Content Analysis*

	<b>Lexical Analysis</b>	<b>Traditional Content Analysis</b>
PAFR	0.00513	18.11111
Budget	0.00523	19.50000
Difference	0.00009	1.38889

Table 4

*Standard Deviation of Budgets and PAFRs*

	<b>1</b>	<b>2</b>	<b>3</b>	<b>Total</b>
PAFR	2.224	2.619	0.000	4.428
Budget	2.673	4.833	0.463	7.616
Difference	0.449	2.214	0.463	3.187

- 1 – Organization of Report
- 2 – Performance Information
- 3 – Communication and Contact

Interestingly, there was less consistency in the scores of budgets compared with the PAFRs. Table 4 (above) shows the standard deviation of budget scores and PAFRs in the traditional content analysis. The high degree of variance in the performance information available in the budgets reinforce the narrative

description offered, above, that not all exemplar state budgets fully integrate performance information into their budgets. Despite the number of states mandating performance budgeting, often the process is not integrated into the budget or it is reported separately. Further, while there is more consistency and less variation among the PAFRs, overall comprehensiveness of these documents is still lacking.

### **CONCLUSION**

Evaluating public performance requires consistent longitudinal metrics. This study builds on existing research that aims to integrate performance measurement with budgeting and reporting efforts in government. Public agencies are continually challenged to measure performance consistently over time, and report the results, which are offered to citizens electronically using the Internet and increasingly social media. States must acknowledge both difficulties and benefits they incur when incorporating performance measurement results into budget and financial reporting processes for the purposes of improving decisions and informing citizens. Integrating performance data into budgeting and financial reporting ensures transparency, improves public trust, and holds agencies accountable. By their nature, fiscal reports explain the past expenditures, while budget reports look forward and represent the intent of public officials for the future. Perhaps the most significant finding in this research is the extent to which performance budgeting is incorporated into exemplar state budgets. This research demonstrates that while performance information is integrated into budgets, the extent to which it is involved in the planning process is less apparent. As statewide performance portals become more popular it is a challenge for states with performance budgeting mandates to include performance data in their budgets. Including performance data in state budgets allows for a historical comparison assessment of agency performance and goal setting for the future. States that fail to integrate performance data in their budgets and instead include a link to a statewide performance portal are in contradiction with

the ideals of performance budgeting. Further they risk violating the spirit, if not the letter, of performance budgeting mandates

The analysis of budget and financial reports through a performance measurement lens suggests that states may find it easier to report on performance as a function of the budget rather than in the context of previous expenditures. This is not to say that states do not address performance in financial reporting, but rather the standards and guidelines for states interested in integrating performance measurement into their budgets are more prevalent as compared to financial reporting. This represents both a weakness and an opportunity, as ghosts of performance budgeting pilots abound to inform current budget officers, while there are fewer examples of popular annual financial reports. Such obstacles should not discourage public administrators from effectively integrating objective performance data into public budget and financial reports on a regular basis.

Integrating performance reporting into budget reports online serves as a mechanism to build public trust. Internally, it serves as a mechanism for holding managers accountable for results and improves effectiveness of public service delivery. For policy makers, the link between performance measurement and policy decisions like budget allocation is less clear. Despite the opaque relationship between performance measurement and policy determinations, the integration of performance information into policy discourse like annual budget deliberations are necessary to inform and ground the process. Including performance measures in budget documents helps public administrators in budgetary decision-making and resource allocation based on the agencies' performance and effectiveness. However, performance measures in budget documents should not just be directed towards the legislative members; they should also be addressed towards citizens in a readable format that enables them to understand how their tax dollars have been spent. Such public reporting can possibly be achieved by states through integrating performance measures into publically available PAFRs, rather than just budget documents. Moreover, performance measures need to be linked to strategic plans to provide a comprehensive perspective of the states' performance.

None of the budgets analyzed in this research integrates strategic performance information across multiple departments. The lack of such an aggregated approach, which is sometimes referred to as a balanced scorecard or dashboard approach, is a weakness of even the best performance budgets.

Moving forward, states need to continue integrating performance information into both their budget and financial reports. In an era defined by decreases in the availability of public funding and in public trust, citizens need more, not less, information about the outcomes associated with both resources spent and also with planned expenditures of public funds. Moreover, this research emphasizes the relevance of GASB suggestions in providing a comprehensive standard for performance reporting among states as well as municipalities. Among those budget and financial reports evaluated, only the Illinois PAFR makes an attempt to follow the GASB suggestions. Although governments are not required to adhere to its standards, it is advised and generally agreed that these standards can produce decision-useful information that both informs the users of government financial reports and aids government officials in providing transparency, accountability, and stewardship evidence. Finally, with the adoption of e-government among state and local governments, there is a renewed expectation among citizen users to obtain relevant and appropriate performance information from government agencies. Based on the findings, many state budget reports included links that directed to performance portals designed specifically for measuring state agency performance. While the GPRA and the reinventing movement revived performance budgeting in the 1990s (Jordan & Hackbart, 1999; Roth, 1992), the combination of e-government and citizen participation has the potential to rejuvenate public performance reporting and strengthen the link between performance measurement and budgeting.

## NOTES

<sup>1</sup> The qualitative research software YoshiKoder was used to conduct the lexical analysis. The software was developed by



Identity Project at Harvard's Weatherhead Center for International Affairs (see [www.yoshikoder.org](http://www.yoshikoder.org) for more details). 22 documents were analyzed to develop the performance dictionary. The analysis of documents yielded a list of over 25,000 distinct terms used in the performance documents. From the list of terms used, the authors each selected 50 key terms that indicates an emphasis on performance or performance measurement. The two lists were compared and a final list of key words was compiled by consensus. The final performance dictionary of 47 terms includes multiple variations of keywords (multiple tenses, plural usage) identified by the authors from the initial performance documents. For example, the terms: performance, perform, performs and performance-measures were all used as distinct terms in the performance dictionary. Using Yoshikoder, frequency and relative frequency (or proportional frequency) of key terms was identified in each of the 9 PAFR and 8 Budget reports.

#### REFERENCES

- Ammons, D. N. (1995). Overcoming the inadequacies of performance measurement in local government: The case of libraries and leisure services. *Public Administration Review*, 55(1), 37-47.
- Anon. (2003). *Reporting performance information: Suggested criteria for effective communication*. GASB-CT. Norwalk.
- Aristigueta, M. P., & Justice, J. B. (2006). Introduction to the symposium: The status of performance budgeting. *Public Performance & Management Review*, 30(1), 7-13.
- Botner, S. B. (1985). The use of budgeting/management tools by state governments. *Public Administration Review*, 45(5), 616-620.

- Caddy, J., & Vergez, C. (2004). *Citizens as partners: Information, consultation and public participation in policy-making*. Paris: Organisation for Economic Co-operation and Development.
- Chadwick, A., & May, C. (2003). Interaction between states and citizens in the age of the Internet: 'E-Government' in the United States, Britain, and the European Union. *Governance*, 16(2), 271–300.
- Connelly, M., & Tompkins, G. (1989). Does performance matter? A study of state budgeting. *Policy Studies Review*, 8(2), 288-299.
- Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, 28(3), 376–417.
- Government Finance Officers Association. (n.d.). Retrieved Oct 25, 2012 from GFOA, [http://www.gfoa.org/index.php?option=com\\_content&task=view&id=34&Itemid=55](http://www.gfoa.org/index.php?option=com_content&task=view&id=34&Itemid=55)
- Gilmour, J., & Lewis, D. (2006). Does performance budgeting work? An examination of the Office of Management and Budget's PART scores. *Public Administration Review*, 66(5), 742-752.
- Grizzle, G. A. (1986). Does budget format really govern the actions of budget makers? *Public Budgeting & Finance*, 6(1), 60–70.
- Ho, A. T.K., & Ni, A. Y. (2005). Have cities shifted to outcome-oriented performance reporting? - A content analysis of city budgets. *Public Budgeting & Finance*, 25 (2), 61–83.

- Holliday, I., & Kwok, R. C. W. (2004). Governance in the information age: Building e-government in Hong Kong. *New Media & Society, 6(4)*, 549-570.
- Holzer, M., Melitski, J., Rho, S. Y., & Schwester, R. (2004). *Restoring Trust in Government: The Potential of Digital Citizen Participation*. E-government Series. Washington, DC: IBM Center for the Business of Government.
- Howard, S. K. (1973). *Changing state budgeting*. Lexington, KY: Council of State Governments.
- Isenmann, R., Bey, C., & Welter, M. (2007). Online reporting for sustainability issues. *Business Strategy and the Environment, 16(7)*, 487-501.
- Jones, D. B., Scott, R. B., Kimbro, L., & Ingram, R. W. (1985). The needs of users of governmental financial reports. Stamford: Government Accounting Standards Board.
- Jordan, M., & Hackbart, M. (1999). Performance budgeting and performance funding in the states: A Status Assessment. *Public Budgeting and Finance, 19(1)*, 68-88.
- Joyce, P. (1993). Using performance measures for federal budgeting: Proposals and prospects. *Public Budgeting and Finance Review, 13(4)*, 3-17.
- Julnes, P. L., & Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review, 61(6)*, 693-708.
- Justice, J. B., Melitski, J., & Smith, D. L. (2006). E-Government as an instrument of fiscal accountability and responsiveness. *The American Review of Public Administration, 36(3)*, 301-22.

- Kahn, J. (1997). *Budgeting democracy: State building and citizenship in America*. Ithaca, NY: Cornell Univ. Press.
- Kasdin, S. (2010). Reinventing reforms: How to improve program management using performance measures. Really. *Public Budgeting & Finance*, 30(3), 51–78.
- Kettl, D. (2000). *The Global Public Management Revolution*. Washington, DC: Brookings Institute Press.
- Lauth, T. (1985). Performance evaluation in the Georgia budgetary process. *Public Budgeting & Finance*, 5(1), 67-82.
- Lee, M. (2004). Public reporting: A neglected aspect of nonprofit accountability. *Nonprofit Management and Leadership*, 15(2), 169-185.
- Lee, M., & IBM Center for the Business of Government. (2004). *E-reporting: Strengthening democratic accountability*. Washington, DC: IBM Center for the Business of Government.
- Lee, R., & Johnson, R. (1994). *Public Budgeting Systems*. Gaithersburg, MD: Aspen Publishers, Inc.
- Lu, Y., Willoughby, K., & Arnett, S. (2009). Legislating results. *Public Performance & Management Review*, 33(2), 266–287.
- MacManus, S. A. (1984). Coping with retrenchment: Why local governments need to restructure their budget document formats. *Public Budgeting & Finance*, 4(3), 58-66.
- Melkers, J. E., & Willoughby, K. G. (2001). Budgeters' views of state performance-budgeting systems: Distinctions across branches. *Public Administration Review*, 61(1), 54–64.

- Melkers, J. E., & Willoughby, K. G. (1998). The state of the states: Performance-based budgeting requirements in 47 out of 50. *Public Administration Review*, 66-73.
- Miles, M., & Huberman, M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage Publications.
- Poister, T. H. (2010). The future of strategic planning in the public sector: Linking strategic management and performance. *Public Administration Review*, 70, 246-254.
- Poister, T. H., & Streib, G. (1999). Performance measurement in municipal government: Assessing the state of the practice. *Public Administration Review*, 59(4), 325-335.
- Roth, W. (1992). Performance-based budgeting to enhance implementation of the CFO Act. *Public Budgeting and Finance*, 12(4), 102-106.
- Rubin, I. (1997). *The politics of public budgeting: Getting and spending, borrowing and balancing* (3rd Ed). Chatham, NJ: Chathan House Publishers.
- Scavo, C., & Shi, Y. (1999). World Wide Web site design and use in public management. *Public information technology: policy and management issues*, 299-330.
- Schick, A. (1966). The road to PPB: The stages of budget reform. *Public Administration Review*, 243-258.
- Wholey, J. S., & Hatry, H. P. (1992). The case for performance monitoring. *Public Administration Review*, 52(6), 604-610.

- Yang, K., & Holzer, M. (2006). The performance–trust link: Implications for performance measurement. *Public Administration Review*, *66(1)*, 114-126.
- Yi, L., & Willoughby, K. (2011). Charting performance reporting: The perspective of state governments. In, 24. Washington, DC, October 15.

## Appendix A

## Articles and Reports used to generate performance dictionary:

- Ammons, D. N. 1995. "Overcoming the Inadequacies of Performance Measurement in Local Government: The Case of Libraries and Leisure Services." *Public Administration Review* 55(1).
- Behn, R. D. 2003. "Why measure performance? Different purposes require different measures." *Public Administration Review* 63(5): 586–606.
- Berman, E., and X. H. Wang. 2000. "Performance measurement in US counties: Capacity for reform." *Public Administration Review* 60(5): 409–420.
- Dubnick, M. 2005. "Accountability and the promise of performance: In search of the mechanisms." *Public Performance & Management Review* 28(3): 376–417.
- Epstein, P. D., and W. Campbell. 2000. *GASB SEA Research Case Study: Iowa*.  
Government Accounting Standards Board-CT, Norwalk.
- Epstein, P. D., and W. Campbell. 2000. *GASB SEA Research Case Study: Louisiana*.  
Government Accounting Standards Board-CT, Norwalk.
- Fountain, J. 2000. *GASB SEA Research Case Study: State of Oregon: A Performance System Based on Benchmarks*. Government Accounting Standards Board-CT, Norwalk.
- Government Accounting Office. 2005. *Performance Measurement and Evaluation: Definitions and Relationships*. Washington, DC: U.S. General Accounting Office, May.
- Hatry, H. P. 1978. "The status of productivity measurement in the public sector." *Public Administration Review* 38(1): 28–33.
- Hatry, H. P. 2006. *Performance measurement: Getting results*. Urban Inst Press.
- Julnes, P. L., and M. Holzer. 2001. "Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation." *Public Administration Review* 61(6): 693–708.

- Kravchuk, R. S., and R. W. Schack. 1996. "Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993." *Public Administration Review* 56(4).
- Melkers, J., and P. Mhatre. 2002. *Case Study: Wisconsin. Use and Effects of Using Performance Measures for Budgeting, Management and Reporting*. Government Accounting Standards Board-CT, Norwalk.
- Melkers, J., and K. Willoughby. 2005. "Models of Performance-Measurement Use in Local Governments: Understanding Budgeting, Communication, and Lasting Effects." *Public Administration Review* 65(2): 180–190.
- Morley, E., S. P. Bryant, and H. P. Hatry. 2001. *Comparative performance measurement*. Urban Inst Press.
- Poister, T. H. 2010. "The Future of Strategic Planning in the Public Sector: Linking Strategic Management and Performance." *Public Administration Review* 70: s246–s254.
- Poister, T. H., and G. Streib. 1999. "Performance measurement in municipal government: Assessing the state of the practice." *Public Administration Review* 59(4).
- Tucker, L. 2000. *GASB SEA Research Case Study: State Of Texas—Focus On Performance*. Government Accounting Standards Board-CT, Norwalk.
- Tucker, L., and W. Campbell. 2002. *Case Study: Maine. Use and Effects of Using Performance Measures for Budgeting, Management and Reporting*. Government Accounting Standards Board-CT, Norwalk.
- Van Thiel, S., and F.L. Leeuw. 2002. "The performance paradox in the public sector." *Public Performance & Management Review* 25(3): 267–281.
- Wholey, J. S., and H. P. Hatry. 1992. "The case for performance monitoring." *Public Administration Review* 52(6): 604–610.
- Yang, K., and M. Holzer. 2006. "The Performance–Trust Link: Implications for Performance Measurement." *Public Administration Review* 66(1): 114–126.



How key words were developed:

22 documents analyzed for key performance related terms. From the list of terms used, the authors each selected 50 key terms that indicates an emphasis on performance or performance measurement. The two lists were compared and a list of key words was generated. The final performance dictionary of 47 terms includes multiple variations of keywords (multiple tenses, plural usage) identified by the authors from the initial performance documents. Using Yoshikoder, frequency and relative frequency of key terms was identified in each of the 9 PAFR and 8 Budget reports.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.